Metaphysical Emergence

Metaphysical Emergence

JESSICA M. WILSON

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX26DP
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2021

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2020943339

ISBN 978–0–19–882374–2

DOI: 10.1093/oso/9780198823742.001.0001

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

For my mother and my husband

List of Figures

Acknowledgments

This book covers research carried out over the past two decades, and I correspondingly have many people and institutions to thank. The heart of the approach to emergence that I endorse here can be traced back to my thesis, *Physicalism, Emergentism, and Fundamental Forces* (Cornell, 2002), prior to that to my first paper, 'How Superduper does a Physicalist Supervenience Need to Be?' (*Philosophical Quarterly*, 1999), and prior to that to John Heil's 1996 NEH seminar in the metaphysics of mind, which Heil graciously allowed me to attend. It was during this seminar that Michael Watkins came up with the 'proper subset of powers' strategy for addressing Jaegwon Kim's concerns that real, distinct, and dependent mental properties would problematically causally overdetermine effects already produced by their physical dependence base properties. I still vividly remember Michael arriving at a cocktail party hosted by co-participant Amie Thomasson and announcing, "I've solved the problem of mental causation!" And so he had, it seems to me, at least from a nonreductive physicalist perspective. Much of my work on physical realization (i.e., Weak emergence) traces back to Watkins's core insight. In the early years developing this approach to realization, along with a 'new power' approach to physically unacceptable (i.e., Strong) emergence, I had the benefit of a dream team committee—Sydney Shoemaker, Richard Boyd, and Jason Stanley—and of supportive interactions with many exceptional philosophers, including Heil, Terry Horgan, and Jaegwon Kim, whose work has been methodologically and dialectically foundational in my own.

For assistance in the ensuing years in developing this project, I am indebted, to start, to my home departments at the University of Michigan (2001–5) and the University of Toronto (2005–present), each of which created environments in which my research was able to flourish. I would like to thank the National Science Foundation for a Professional Development Fellowship providing me with a year off to gain further education in physics, and the Social Science and Humanities Research Council for fellowship support for the project 'Metaphysical Emergence: The Ontological and Causal Autonomy of Special-Science Entities', which resulted in a first draft of this book, and the in-progress project 'How Metaphysical Dependence Works', which has seen the book's completion. I am also grateful to several departments which hosted me as a visiter for extended periods, including the Centre for Consciousness, Philosophy Program, Research School of Social Sciences, Australian National University (as Visiting Fellow), the Logic, Language, and Cognition Research Group at the University of Barcelona (as Visiting Scholar), the Arché Centre for Metaphysics, Language, Logic, and

Epistemology at Saint Andrews (as Visiting Fellow), the Department of Philosophy at the University of Cologne (as Visiting Fellow), the Eidyn Research Centre at the University of Edinburgh (as Regular Distinguished Visiting Professor), and the Institute for Philosophical Research at the National Autonomous University of Mexico (as Visiting Fellow). My experiences and interactions with the many helpful and intellectually incisive philosophers at my home and other institutions have been of enormous benefit to the development of my research on emergence.

I have given many dozens of talks at colloquia, conferences, and workshops on topics related, one way or other, to the present work—too many to list here. For stimulating feedback and conversation I thank the organizers of and audiences at these events. I am also indebted to certain individuals, in addition to those previously mentioned, for valuable assistance in refining aspects of this work, one way or another. These include Ken Aizawa, Torin Alter, Fatema Amijee, Louise Antony, Joseph Baltimore, Umut Baysan, George Bealer, Mark Bedau, Karen Bennett, Sara Bernstein, Ricki Bliss, Stephen Biggs, Alexander Bird, Amanda Bryant, Claudio Calosi, Alex Carruth, David Chalmers, Jonas Christensen, Lenny Clapp, Troy Cross, Louis deRosset, Ranpal Dosanjh, Douglas Ehring, Kit Fine, Steven French, Sophie Gibb, Carl Gillett, Matthew Haug, Robin Hendry, Eric Hiddleston, Robert Howell, Paul Humphreys, Gordon Kane, Kathrin Koslicki, David Kovacs, Maurice Lamb, Tom Lancaster, Pat Lewtas, Cynthia and Graham Macdonald, Cristian Mariani, Jonathan McCoy, Michael McKenna, Brian McLaughlin, Andrew Melnyk, Sandra Mitchell, Barbara Montero, Margie Morrison, Adam Murray, Alyssa Ney, Daniel Nolan, Paul Noordhof, Tim O'Connor, Asya Passinsky, Laurie Paul, Derk Pereboom, Thomas Polger, Diana Raffman, Gonzalo Rodriguez-Pereyra, Luke Roelofs, Jonathan Schaffer, Christopher Shields, Ted Sider, Michael Silberstein, Daniel Stoljar, Michael Strevens, Tuomas Tahko, Elanor Taylor, Mariam Thalos, Alessandro Torza, Bob van Gulick, Jennifer Wang, Gene Witmer, Stephen Yablo, David Yates, and Jeffrey Yoshimi. I thank each of these philosophers for their contributions and for taking the time to engage with me and/or my work, and I apologize to anyone I have inadvertently left off this list. I would like to extend special thanks to Catharine Diehl, Kevin Morris, Michele Paolini Paoletti, and Bill Seager, as well as two anonymous referees for Oxford University Press, for detailed comments on previous drafts; this book was greatly improved by taking these comments into consideration. Thanks as well to Peter Momtchiloff, the paradigm of a patient and encouraging editor, and to others at OUP, including Henry Clarke and Vaishnavi Ananthasubramanyam, who so helpfully assisted in the production of this book.

Some of the material in this book is based on previously published work. For permission to reproduce the occasional short passage I'd like to thank Brill ('Metaphysical Emergence: Weak and Strong', in *Metaphysics in Contemporary Physics*), Cambridge University Press ('Free Will and Mental Quausation', in *Journal of the American Philosophical Association*), Oxford University Press ('Non-reductive Physicalism and Degrees of Freedom', in *The British Journal for the*

Philosophy of Science; 'Non-Reductive Realization and the Powers-based Subset Strategy', in *The Monist*), and Springer Publishing ('The A Priority of Abduction' and 'Determination, Realization, and Mental Causation', each in *Philosophical Studies*). Full citations can be found in the bibliography.

Finally, I want to thank two members of my family, to whom this book is dedicated. Thanks to my wonderful mother, Marilyn Wilson, for being a constant source of love, wisdom, and inspiration. And thanks to my beloved husband, Benj Hellie, for being the most brilliant interlocuter and partner I could ever hope for, in philosophy and in life.

1
Key issues and questions

1.1 Metaphysical emergence: dependence with autonomy

1.1.1 Dependence with autonomy

Consider some entities (objects, systems, or other particulars) of the sort characteristic of the special sciences: cells (treated by cellular biology), organs (treated by biology), trees (treated by botany), birds (treated by zoology), lakes and mountains (treated by geology), hurricanes (treated by meteorology), and humans (treated by anthropology and psychology, among other disciplines), to name just a few. Such macro-entities, the scientists tell us, are ultimately dependent on complex configurations of fundamental physical entities—'micro-configurations', for short—in that, at any given time, a macro-entity inherits its matter from some micro-configuration at that time, and a macro-entity's features (its states, properties, behaviors, or other ways to be) over a given temporal interval are at least in part a function of features of its underlying micro-configuration(s) over that temporal interval. Notwithstanding this cotemporal material dependence, however, these special-science entities also seem, from both theoretical and experiential points of view (to be discussed in more detail shortly), to possess a certain degree of ontological and causal autonomy—that is, they appear to be *distinct* from, and *distinctively efficacious* with respect to, the micro-configurations upon which they depend.

It is the coupling of *cotemporal material dependence* with *ontological and causal autonomy* which is most basically definitive of the notion of emergence, at least as suggested by the central cases of special-science entities with respect to the physical micro-configurations which are their constant companions. (Note that the dependence at issue here need not be instantaneous, but is broad in perhaps holding over a temporal interval. Note also that the notions of autonomy at issue here follow common use in indicating that some goings-on are distinctive in a specific way, leaving open, as in developmental, political, and other contexts, that autonomous goings-on are not *entirely* independent of other goings-on.) This general notion of emergence is also motivated by attention to cases of artifacts, including tables, paintings, buildings, and so on; for these too are reasonably taken to be cotemporally materially dependent on, yet distinct from and distinctively efficacious with respect to, underlying micro-configurations. And insofar as the

Metaphysical Emergence. Jessica M. Wilson, Oxford University Press (2021). © Jessica M. Wilson.
DOI: 10.1093/oso/9780198823742.003.0001

dependence, distinctness, and distinctive efficacy seemingly at issue in all these cases appear to be characteristic of the entities themselves, as opposed to what we can know or represent about such entities, the notion of emergence initially motivated by these cases is that of specifically *metaphysical* emergence.

As we'll see, there are questions about how to interpret the appearances of metaphysical emergence, and relatedly, about whether these appearances can be taken at realistic face value. Such questions will occupy much of this book. First, though, it's worth saying a bit more about the prima facie motivations for there being metaphysical emergence, to substantiate that there is good reason to attend to the content, viability, and applicability of this interesting notion.

1.1.2 The prima facie motivations

Cotemporal material dependence

Why think that special-scientific and artifactual macro-entities are cotemporally materially dependent on micro-configurations of fundamental physical entities—that is, on collections ('pluralities') or structural aggregates of entities of the sort treated by fundamental physics?[1] The motivations here are broadly empirical, and reflect common scientific consensus on two points. The first is that the only matter or substance is physical matter or substance, such that any and all macro-entities (objects, systems, or other particulars) inherit their matter from (configurations of) their ultimately physical constituents. The second is that the features (states, properties, behaviors, or other ways for a particular to be) of any macro-entity over a given temporal interval are at least in part a function of the features of the micro-configuration(s) which materially constitute the macro-entity during that temporal interval.

The introductory section of practically any scientific textbook is likely to register one or both of these baseline scientific assumptions, as in these examples, drawn at random from my bookshelf:

> *Everything is made of atoms*. That is the key hypothesis.
>
> (Feynman 1963, Vol I, 1–9)
>
> Quarks and leptons are the fundamental objects of which all matter is composed; they interact via the exchange of gauge bosons. (Kane 1993, 1)
>
> In the present state of scientific knowledge, quantum mechanics plays a fundamental role in the description and understanding of natural phenomena.

[1] Here and throughout I follow physicists and others in being generous in my use of the expression 'fundamental' when speaking of physical goings-on of the sort treated by fundamental physics. Nothing deep for present purposes turns, for example, on whether physical particles such as electrons are nonfundamental waves in more fundamental physical fields. I say more about the operative notion of fundamentality as relevant to the present project in §1.4.3.

> [...] when we are concerned only with macroscopic physical objects [...] it is necessary, in principle, to begin by studying the behaviour of their various constituent atoms, ions, electrons, in order to arrive at a complete scientific description. There are many phenomena which reveal, on a macroscopic scale, the quantum behaviour of nature. It is in this sense that it can be said that quantum mechanics is the basis of our present understanding of all natural phenomena, including those traditionally treated in chemistry, biology, etc.
> (Cohen-Tannoudji *et al.* 1977, 9)

> Every day, whether we know it or not, we witness changes in matter that are a result of the properties of the atoms and molecules composing that matter—ice melts, iron rusts, gasoline burns, fruit ripens, water evaporates. [...] If we want to understand the substances around us, we must understand the physical and chemical properties of the atoms and molecules that compose them—this is the central goal of chemistry. (Tro *et al.* 2016, 2)

These introductory gestures leave a lot open, but the core idea that macro-entities are cotemporally materially dependent on configurations of physical entities is about as uncontroversial as it gets in the sciences. That said, we will see that there is controversy about exactly how to understand the form(s) of dependence at issue here; moreover, we will also later discuss views, such as substance dualism and panpsychism, which reject the appearances of cotemporal material dependence, on at least some understandings. At this point we are simply canvassing prima facie reasons to think that there is cotemporal material dependence in the cases at issue; and there isn't any real question about there being an enormous amount of empirical support for such a view.

Ontological and causal autonomy

Next, why think that special scientific and artifactual macro-entities are autonomous, ontologically and causally, with respect to the micro-configurations upon which they cotemporally materially depend? Three prima facie motivations for one or both forms of autonomy have to do with how special-science entities are classified or characterized, as follows:

- *Distinctive special-scientific taxonomy*: Special-science entities are classified as falling under distinctive types. For example, gases, cells, hurricanes, trees, birds, and humans are classified as falling, respectively, under distinctive types in, respectively, thermodynamics, biology, meteorology, botany, zoology, and anthropology (as well as psychology and other disciplines); and similarly for countless other macro-entities treated by these and other special sciences. On the face of it, such special-science types are different from those under which (configurations of) fundamental physical entities fall.

Classification practices thus provide prima facie support for thinking that special-science entities are ontologically autonomous with respect to—that is, distinct from—their underlying micro-configurations.

- *Distinctive special-science features*: Special-science entities are characterized as having distinctive features, constitutive of the distinctive types under which they fall. A tree, for example, has roots, a trunk, branches, stems, leaves; it obtains nutrients from air, sun, soil, and water through leaves and roots; it reproduces via seeds and may bear fruit; it is deciduous or evergreen; it is hardy in certain climate zones, and so on. On the face of it, such features are not appropriately attributed to even complex configurations of fundamental physical entities; and the same is true for the characteristic features of other special-science entities. The mental features of human persons are especially distinctive, as involving, among other salient characteristics, qualitative aspects of sensory experience; representational aspects of belief, desire, intention, and other mental states; appreciation of aesthetic, moral, and other broadly normative values; and seemingly free agency. By Leibniz's Law, of course, entities with different features (at the same time, at least) are distinct; hence that special-science entities are characterized as having features different from those had by the micro-configurations upon which they cotemporally materially depend again supports special-science entities' being ontologically autonomous with respect to these micro-configurations.
- *Distinctive special-science laws*: Special-science entities are taken to be governed by special-science laws (or regularities) describing states, properties, and behaviors of, including associated causal interactions involving, such entities—laws that, on the face of it, are different from those governing even complex configurations of fundamental physical entities. To be sure, there are deep questions here (which will be a primary focus in what follows) about how to understand the distinctive efficacy of special-science goings-on, and relatedly, about whether such efficacy is compatible with the assumed efficacy of micro-configurations and constituent fundamental physical entities.[2] At present, we can say this much: that special-science laws are seemingly distinctive and seemingly causal provides prima facie support for special-science entities' being causally autonomous—that is, distinctively efficacious—with respect to their underlying micro-configurations.

Two other reasons for thinking that special-science entities are autonomous reflect, more specifically, that special-science entities, features, and laws often

[2] It is sometimes claimed, typically by appeal to the discussion in Russell 1912, that physics has dispensed with the notion of 'cause', but these complaints, old and new, are directed at specific and sometimes implausible conceptions of causation, leaving open the possibility of broader and empirically informed conceptions of physical causation. I revisit the issue of physical causality in §1.4.4.

abstract from details relevant to characterizing physical (more generally: dependence base) entities, features, and laws:

- *Universal properties and behavior*: Many special-science entities, including thermodynamic complex systems such as liquids and gases, exhibit features that are functionally independent of various features of their underlying micro-configurations. For example, complex systems near critical points exhibit 'universal' properties and behavior across widely diverse underlying micro-configurations. Such cases provide support for thinking that some special-science entities are causally, hence ontologically, autonomous, in that their behaviors are sensitive to comparatively abstract causal joints or levels of grain.
- *Elimination of microphysical 'degrees of freedom'*: Degrees of freedom are independent parameters needed to specify states relevant to an entity's law-governed properties and behaviors. In some cases, certain such states of special-science entities (e.g., the configuration state, encoding an object's or system's spatial location) are specifiable by reference to strictly fewer degrees of freedom than are needed to specify the corresponding (e.g., configuration) state of the micro-configurations upon which the special-science entities cotemporally materially depend. For example, rigid bodies cotemporally materially depend upon quantum-mechanical micro-configurations, but the behavior of rigid bodies does not generally rely upon quantum-mechanical degrees of freedom such as spin direction and magnitude, as is reflected in the fact that the laws governing rigid bodies do not contain reference to such degrees of freedom (or their values). That certain law-governed features of rigid bodies can be specified using strictly fewer degrees of freedom than are required to specify the associated law-governed features of their underlying quantum-mechanical micro-configurations suggests that the former are ontologically autonomous with respect to the latter.

Four other prima facie motivations for autonomy apply to artifacts of the sort available to ordinary experience as well as certain special-scientific entities, and reflect perceptual, individuative, and semantic considerations, in ways that support the scientific characterizations of such entities as having features which abstract from certain details of the micro-configurations upon which they depend:

- *Perceptual unity*: Though the macro-entities of our acquaintance are, scientists tell us, materially constituted by massively complex and constantly changing micro-configurations, macro-entities do not perceptually appear to us as massively complex, constantly changing, configurations of micro-entities. A tree, for example, does not perceptually appear as a complicated structure of cells or tissues, much less as a buzzing array of subatomic

particles or other physical fundamenta; rather, a tree perceptually appears as a comparatively stable and unified entity, both at and over time; and the same is true for statues and other familiar macro-entities. Moreover, in the case of human persons there is the evidence of introspection—a kind of internal perception—of ourselves as fairly unified and persisting 'selves'.

- *Compositional flexibility*: The identity of many macro-entities appears to transcend that of their underlying micro-configurations, in the sense that the existence of a given special-science entity (a hurricane, a cell, a tree, a bird, a human) does not depend on the existence of any *specific* micro-configuration(s), and relatedly, in that many macro-entities, both in scientific practice and in our ordinary practices of individuation, are taken to be capable of surviving at least some changes in their underlying micro-configuration(s)—as when, e.g., the Venus de Milo survived the loss of her arms.
- *Proper names and definite descriptions*: Our practices of giving names or definite descriptions to certain artifacts and special-science entities—the Mona Lisa, the CN Tower, the Atlantic Ocean, Hurricane Katrina, Benj Hellie—suggests that we often treat macro-entities as individuals, distinct from the ever-changing micro-configurations upon which they cotemporally materially depend.
- *Truth and meaning*: Many of the sentences that we take to be true appear to be about macroscopic goings-on, and relatedly, to contain subject terms which appear to denote macro-entities (as either token individuals or types) and predicate terms which appear to be used to attribute macro-features to these macro-entities—as with, e.g., 'That table is well-made', 'Tigers typically have tails', and 'Her novels are spellbinding'.

One final motivation for ontological and causal autonomy is worth mentioning, as especially dear to our hearts and minds:

- *Seemingly free will*: The status of our actions as genuinely free remains up for debate. However, a starting point in this debate is that it at least introspectively seems as if creatures like ourselves are capable of making free choices to produce (or intend to produce) certain effects, where this efficacy appears to be quite different from that associated with the (deterministically or indeterministically) lawfully governed micro-configurations upon which we and our mental states cotemporally materially depend.

Summing up: many considerations, drawn from science, perception, language, our practices of individuation, and introspective experience, provide prima facie support for thinking that many broadly natural entities are *cotemporally materially dependent* on micro-configurations of fundamental physical entities, yet are also

ontologically and causally autonomous with respect to these underlying micro-configurations. And that is just to say that, on the face of it, there is metaphysical emergence.

1.2 Two key questions

Given that the prima facie motivations for metaphysical emergence span such a wide range of phenomena, the interest in exploring and illuminating this notion is clear. Towards this end, two questions (and related sub-questions) are key.

The first pertains to the nature and varieties of metaphysical emergence. Here we ask: just what is metaphysical emergence, more precisely? How is it, exactly, that special scientific and artifactual entities can cotemporally materially depend on complex configurations of fundamental physical entities, while retaining some degree of ontological and causal autonomy? And is there more than one way in which this can be—is there more than one form of metaphysical emergence?

The second pertains to whether there actually is any metaphysical emergence. To start: are there any insuperable problems with the notion(s) of metaphysical emergence, such that emergence is, at best, an epistemic or representational phenomenon? And supposing that one or more varieties of metaphysical emergence is in-principle viable, are there any actual cases of such emergence? If we can make sense of metaphysical emergence, both in principle and as actually instantiated, then we would be on track to vindicating and illuminating the appearances of emergence in science and ordinary experience. But if we cannot make sense of emergence as viable or as actual, that would still be worth knowing. Either way, the result promises to have widespread ramifications for our understanding of the world around us.

There is good reason, then, to pursue answers to the key questions of what, more specifically, metaphysical emergence is, and whether there actually is any. Indeed, in past decades there has been an explosion of interest in emergence, in both philosophical and scientific contexts. Unfortunately, as I'll now discuss, all this attention has left the answers to the key questions less clear than ever.

1.2.1 What, more specifically, is metaphysical emergence?

One source of unclarity as regards the question 'What is metaphysical emergence?' stems from the availability of multiple interpretations of the notions of dependence and autonomy at issue, which interpretive options have given rise to a huge diversity of accounts offered as providing comparatively specific answers to this question. Before surveying this diversity, however, it's worth first registering

certain ways in which accounts of metaphysical emergence directed at the target cases typically agree, since these commonly held assumptions will be largely taken for granted in what follows.[3]

The main points of agreement are as follows:

- Accounts of metaphysical emergence typically agree in taking the only substance to be material or physical substance, and relatedly, in taking emergence not to involve any new substance, of the sort posited, e.g., by Cartesian dualists or vitalists.[4] As Stephan (2002) puts it:

  > The first feature of contemporary theories of emergence, the thesis of physical monism, is a thesis about the nature of systems that have emergent properties (or structures). The thesis says that the bearers of emergent properties are made up of material parts only. It denies that there are any supernatural components responsible for a system's having emergent properties. Thus, all substance-dualistic positions are rejected [...]. (79)

  Hence, for example, Mill (1843/1973), the father of British Emergentism, supposes that emergent entities are entirely materially composed:

  > All organised [living] bodies are composed of parts similar to those composing inorganic nature, and which have even themselves existed in an inorganic state; but the phenomena of life which result from the juxtaposition of those parts in a certain manner bear no analogy to any of the effects which would be produced by the action of the component substances considered as mere physical agents. (243)

[3] As regards the identification of these common components, it is worth appreciating that not every account or treatment of so-called 'emergence' is relevant to investigations into emergence of the sort motivated by the target cases of special-scientific and artifactual entities and features. For example, the diachronic conception of emergence at issue in Morgan's (1923) notion of 'emergent evolution', whereby emergence reflects increasingly complex combinations of lower-level constituents, is not relevant to making sense of emergence in the target cases. Perhaps relatedly, nor are 'part/whole' conceptions on which it suffices for emergence that a whole has a feature not had by any individual part (see §1.4.2 for further discussion).

Also not to the point of accommodation of the target cases is the sort of intra-level diachronic emergence advanced in Humphreys (1997), involving the exhaustive (non-mereological) 'fusion' of multiple physical entities into another physical entity (see also the 'horizontal' account of emergence in Santos 2015 and accounts of 'transformational emergence' in Humphreys 2016 and Guay and Sartenaer 2016). More generally, I am inclined to think that such cases of diachronic intra-level 'emergence' can be understood simply as involving causation (see Wilson in progress*a*); hence I do not follow Paolini Paoletti (2018*b*) in taking it to be an advantage of an account of emergence that it be general enough to accommodate such cases.

[4] As in Stephan's remarks to follow, 'material' and 'physical' are sometimes treated as synonyms in the contemporary literature, at least so far as characterizing substance is concerned. I discuss the operative notion of the physical in §1.4.1; for present purposes the supposition of physical monism can be understood as the view that there is only one kind of substance, of the sort that was traditionally characterized as 'matter' or 'material' substance.

Moreover, on contemporary accounts of emergence, it is also typically supposed that what it is for the entities constituting the 'stuff' of natural reality to be physical is for them to be among the target subject matter of fundamental physics.[5] In what follows, I will also suppose that all substance is physical substance, and that what this comes to is that the compositionally basic entities are physical.[6] That said, accounts of emergence typically admit of generalization beyond the specific commitment to material or physical substance monism, in the sense that they could be applied to other forms of substance monism.

- Accounts of emergence typically agree that core to the notion is the combination of cotemporal material dependence and (some degree of) autonomy. These core components are sometimes explicitly flagged (see Bedau 1997), and are sometimes implicitly encoded in specific accounts of the dependence and autonomy at issue, as when Kim (2006) says

  > [T]wo [...] necessary components of any concept of emergence that is true to its historical origins [...] are supervenience and irreducibility. (548)

  As previously noted, in accounts of metaphysical emergence the intended notion of cotemporal dependence is broad, in applying either at a time or over a temporal interval—i.e., is one allowing for temporally extended base and dependent phenomena. And in these accounts the intended notion of autonomy is one entailing not just that emergents are distinct (ontologically autonomous) from their dependence bases, but also that emergents are distinctively efficacious (causally autonomous) with respect to these bases. To be sure, even as regards certain of these components there is some dispute; such variations, however, are either subsumable under the core understandings or else not to the point of accommodating the target phenomena involving special-scientific or artifactual entities (or features) and their material bases.[7]

[5] As Paolini Paoletti (2018*b*) puts it, "Microphysical entities are here taken to be all the relatively compositionally basic entities in the universe, as they are studied (or they can be studied) by physics" (4).

[6] What if there are no compositionally basic entities? I'll address this issue in §1.4.1, when I discuss the operative notion of the physical.

[7] Again, not every account of so-called 'emergence' is directed at the target cases of special-scientific and artifactual entities and features; the accounts discussed in this note, however, are directed at these cases.

In re purportedly diachronic emergence: Mill (1843/1973) suggests that certain ("heteropathic") effects diachronically emerge from temporally prior causes, but as he observes, one can understand such emergence as involving the cotemporal emergence of powers of a composite entity to produce such effects. A similar treatment accommodates the view of O'Connor and Wong (2005), according to which emergent features are diachronically caused by dependence base features. More generally, since whether an entity causes a given effect typically requires that certain circumstances be in place, it is preferable to treat 'causal' accounts of emergence in terms of the emergence of powers of a composite;

- Accounts of metaphysical emergence typically agree that the emergence of entities can be investigated by attention to the emergence of features of the entities at issue. On this understanding, any emergence there might be involves an emergent feature. As Bedau (2002) puts it:

  > [A]n entity with an emergent property is an emergent entity and an emergent phenomenon involves an emergent entity possessing an emergent property—and they all can be traced back to the notion of an emergent property. (6)

  Here the supposition is that if some entity is emergent, this is because it has some characteristic feature which is emergent—e.g., the feature (characteristic of a liquid or gas near a critical point) of *undergoing a change of phase*, or the feature (of a human or other minded creature) of *being in pain*—which can be the direct target of investigation. Relatedly, a focus on features is useful in that talk of entities can be naturally translated into talk of features—namely, the property or feature of being an entity of the type in question.[8]

  I too will usually (though not always) focus on the emergence of features as offering a comparatively straightforward and systematic way of investigating the emergence of associated entities.[9] Three clarificatory points in re this focus are in order. First, insofar as emergence involves cotemporal material dependence, which in the first instance applies to entities, we need to extend the operative understanding of cotemporal material dependence to apply to features, as follows:

for otherwise the emergence of the composite would implausibly depend on whether it happened to occupy circumstances enabling it to enter into the production of the effect in question.

Another supposed motivation for a diachronic notion of emergence is that emergence can be temporally extended, as with, e.g., Rueger's (2000) account of diachronic emergence as involving temporally extended processes, and Mitchell's (2012) account of emergence as involving 'dynamic self-organization' (see also Kirchhoff 2014). But again, the emergence of such processes is compatible with these cotemporally depending on a temporally extended base; compare Lepore and Loewer's (1989) discussion of how, even if externally individuated content properties do not supervene on spatiotemporally local (i.e., neurophysiological) properties, "there may be a more global physical property whose instantiation explains the possession of content properties" (181).

In re purportedly epiphenomenal emergence: Chalmers (1996) and Morris (2014) characterize an epiphenomenalist form of emergence, on which emergents are ontologically but not causally autonomous. However, such emergence is beside the point of accommodating the emergence of special-science entities as entering into distinctive, typically causal, laws, and so (modulo discussion in Ch. 4, §4.4.1) will be put aside here.

[8] A third reason to focus on the emergence of features is that, and notwithstanding that the initial motivations for metaphysical emergence are naturally expressed as involving the emergence of entities, one might want to leave open the possibility that one feature of an entity might emerge from another feature of that same entity. In Ch. 4 (§4.2.3) I will revisit the question of whether the emergence of a feature always brings the emergence of an entity in its wake, and if so, whether this is in tension with the assumption of physical monism.

[9] Nominalists who reject the existence of properties or other features are invited to apply their preferred strategies for converting talk of features into talk of objects or other entities in what follows.

> What it is for a feature *S* to cotemporally materially depend on a feature *P* is for the entity bearing *S* to cotemporally materially depend on the entity bearing *P*.[10]

Second, sometimes the emergence of the *type* of feature (entity) will be at issue, where a type of feature may have many instances or tokens; other times the emergence of a given *token* of a feature (entity) will be at issue. When the difference between type and token matters, this will be made clear. Third, while most accounts of metaphysical emergence take this to involve a one-one relation between base and emergent features, some accounts characterize the relation between features as a many-one relation, with a given feature emerging from a plurality of features or feature instances. I am inclined to think that many-one approaches to metaphysical emergence can either be subsumed under a one-one approach or else are not to the point of characterizing emergence understood as an inter-level relation,[11] but I will track the difference where it matters.

- Accounts of metaphysical emergence typically agree that emergence has certain correlational or modal implications, sometimes expressed by saying that emergent features 'supervene' on base features with at least 'nomological'—i.e., natural law-based—necessity. As I'll sometimes put it: emergent features 'minimally nomologically supervene' on base features.[12] Here the idea is that in every world (actual or hypothetical) with the same or relevantly similar laws of nature, the occurrence of an emergent feature *S* requires the occurrence of some or other base feature *P*, and in every such world, the occurrence of any such *P* will be accompanied by the occurrence of such an *S*. For example, Broad (1925) maintains that emergent features of a compound are "completely determined" by features of its parts when appropriately configured, in that "whenever you have a whole composed of these [...]

[10] If one wants to make room for the possibility that a feature of an entity can emerge from that same entity (as per note 8), one should allow here that an entity can cotemporally materially depend on itself.

[11] I'll revisit this issue in Ch. 3 (§3.4.2), when discussing the 'many-one' approach endorsed in Gillett 2010, 2002*a*, and 2016. On Gillett's approach, emergence involves the cotemporal material dependence of a single (e.g., special-scientific) feature on multiple, comparatively nonrelational (e.g., physical) features, where the latter are understood as combining in various (e.g., lower-level causal, spatiotemporal, or mereological) ways agreed by all parties not to involve emergence, whereas the one-one approach takes emergence to involve the cotemporal material dependence of a single (e.g., special-scientific) feature on a single (e.g., physical) feature of a micro-configuration (plurality or structural aggregate), understood as already combined in various ways agreed by all parties not to involve emergence. So understood, the many-one approach is subsumable under the one-one approach; see related discussion in Bennett 2017, 10. One reason to stick with a one-one approach is that intra-level causation may also take the form of a many-one relation; but such causation is beside the point of emergence as an inter-level relation.

[12] See Kim 1990 and McLaughlin and Bennett 2018 for discussion of the notion and varieties of supervenience.

elements in certain proportions and relations you have something with the [compound's] characteristic properties" (64).

As we'll see, some accounts of metaphysical emergence take the correlations to hold with just nomological necessity, whereas others take them to moreover hold with metaphysical necessity (that is, in every possible world—not just worlds with the same or relevantly similar laws of nature). Either way, these accounts all agree that emergent features minimally nomologically supervene on base features.

- Finally, accounts of metaphysical emergence typically agree that natural reality exhibits a kind of 'leveled' or 'layered' structure, reflected in the broadly hierarchical structure of the special sciences. Certain views contrasting with any form of emergentism—e.g., vitalism or substance dualism—also take natural reality to have a leveled structure, so the supposition that there are multiple levels is not equivalent to the supposition that there is emergence: accounts of metaphysical emergence differ from these other views in aiming to make sense of natural reality's having a leveled structure in a way compatible with the appearances of cotemporal material dependence coupled with autonomy. That qualification registered, emergent entities and features are often characterized as 'higher-level' with respect to the 'lower-level', and ultimately physical, goings-on upon which they depend. Talk of levels is convenient, especially as making room for there being different yet concurrently existing systems of entities, features, and laws of the sort that are associated with the special sciences or different scales of experience or reality, and in what follows this terminology will frequently make an appearance (as in, e.g., discussion of Kim's 'problem of higher-level causation').

  There remains the independent and important question of how to individuate levels in a way that does not prejudice the question of whether there is genuine metaphysical emergence against the ontological or metaphysical reductionist. The reductionist maintains, contra any metaphysical emergentist, that all goings-on, including apparently 'higher-level' goings-on, are in fact identical to some or other complex combination of 'lower-level', ultimately physical goings-on. Correspondingly, the operative conception of levels needs to be broad enough to not immediately rule out of court the reductionist view that there is only one level, such that, e.g., the level of physical goings-on that all parties agree (in the present debate) contains the compositionally basic entities needs to include not just, e.g., physical particles such as quarks and electrons standing in relatively noncomplex physical relations, but also massively complex pluralities or structural aggregates of such particles and relations (where the relations at issue include lower-level causal, spatiotemporal, mereological, and other ontologically 'lightweight' modes of combination), of the sort that might potentially serve as the

> candidate physicalist reduction base for macro-entities such as tables, planets, and persons, and their characteristic features. I'll return to the question of how to individuate levels in §1.4.2. Meanwhile, one should keep in mind that the notion of a 'level' in what follows must be suitably expansive if substantive debate between the reductionist and emergentist is to proceed.[13]

So there are important core points of agreement among accounts of metaphysical emergence. Beyond these core points of agreement, however, accounts of such emergence diverge into a bewildering variety, primarily reflecting that the core notions of dependence and autonomy have multiple, often incompatible interpretations. The extent of this diversity has led some to claim that references to emergence "seem to have no settled meaning" (Byrne 1994, 206); that accounts of emergence are "not obviously reconcilable with one another" (O'Connor 1994, 91); that there has been "a historical load of confusion surrounding the metaphysical aspects of the concept, reflected in the fact that it has been used in a long series of different ways" (Emmeche *et al.* 1997, 84); that "those discussing emergence, even face to face, more often than not talk past each other" (Kim 2006, 548); that "'emergent' and all its semantic kin have come to stand for a hopeless jumble of different ideas" (Ladyman and Ross 2007, 193); and that "within philosophy and the sciences the term 'emergence' is used in such a bewildering variety of ways that it seems the word itself is the only thing shared across these various usages" (Silberstein 2009, 254).

To be sure, some of the confusing diversity to which these philosophers are adverting reflects that some accounts of emergence take this to be a merely epistemic or representational phenomenon. On these other approaches, the seeming autonomy of emergent phenomena is understood not in metaphysical terms, but rather in terms of such phenomena's being, e.g., unpredictable or underivable from lower-level theories, and where the failures of predictability or derivability are not taken to have any clear metaphysical consequences for whether there are distinct and distinctively efficacious higher-level entities.[14] But even restricting ourselves

[13] It may also be worth registering that there need be no supposition that natural (or artifactual) reality neatly divides into anything resembling a layer cake. The complete characterization of creatures like us, for example, would presumably draw upon goings-on at multiple (physical, chemical, biological, psychological, ecological) levels. This complexity provides another reason why a focus on the emergence of features rather than entities may in some cases be more perspicuous.

[14] Discussions of emergence in scientific contexts are commonly cashed in (purely) epistemic terms. Hence, for example, Anderson (1972) qualifies the metaphysical import of his discussion of the unpredictability-based emergence of certain features of an ammonia molecule by saying that "we must all start with reductionism, which I fully accept" (22). See also, e.g., the overview of scientific accounts of emergence in terms of unpredictability or surprise in Damper 2000. Such purely epistemic accounts are for the most part irrelevant to the present project.

to accounts of emergence intended to have such metaphysical consequences, there remains a bewildering variety of options.[15]

Candidate accounts of the dependence at issue in metaphysical emergence include mereological ('part/whole') determination,[16] causation or nomological connection,[17] functional realization,[18] constitutive mechanism,[19] the determinable/determinate relation,[20] inheritance of causal powers,[21] and primitive 'Grounding'.[22] Candidate accounts of the ontological and/or causal autonomy at issue are even more various. Explicitly metaphysical accounts of such autonomy include nomological but not metaphysical supervenience,[23] nonfundamental novelty (of features, powers, laws, entities),[24] fundamental novelty (of features, powers, forces/interactions, laws, entities),[25] nonadditivity/nonlinearity,[26] 'downward' causal efficacy,[27] multiple realizability/universality/compositional plasticity,[28] causal proportionality/difference-making/counterfactual considerations,[29] elimination in degrees of freedom,[30] sometimes associated with symmetry breaking,[31] and the holding of a proper subset relation between token powers,[32] sometimes cashed in terms of a proper parthood relation between properties or behaviors.[33] And 'epistemic criteria' accounts of ontological and/or causal autonomy include in-principle failure of deducibility/predictability/explicability,[34] predictability, but only by simulation,[35] lack of conceptual or representational entailment,[36] and theoretical/mathematical singularities.[37] Given this plethora

[15] The lists and citations to follow are representative rather than exhaustive; there are many hundreds of papers and books on these notions and their variations, as entering into accounts of either physically unacceptable emergence (a.k.a. 'strong' emergence) or physically acceptable emergence (a.k.a. 'realization' or 'weak' emergence).

[16] See Stephan 2002, Gillett 2002*a*.

[17] See Searle 1992, O'Connor and Wong 2005.

[18] See Putnam 1967, Boyd 1980, Poland 1994, Antony and Levine 1997, Melnyk 2003, Yates 2012.

[19] See Craver 2001, Haug 2010, Gillett 2016.

[20] See Macdonald and Macdonald 1986, Yablo 1992, Ehring 1996, Wilson 2009.

[21] See Kim 1992*a*, Wilson 1999 and 2015*b*, Shoemaker 2000/2001.

[22] See Schaffer 2009, Dasgupta 2014.

[23] See van Cleve 1990, Chalmers 1999, Seager 1999/2016, Noordhof 2010.

[24] See Humphreys 1996, Wimsatt 1996, Crane 2001, Pereboom 2002, Megill 2013.

[25] See the British Emergentists (e.g., Mill 1843/1973, Alexander 1920, Broad 1925), Kim 1992*a*, Cunningham 2001, O'Connor 1994, Wilson 2002*a* and 2015*b*, Barnes 2012, Paolini Paoletti 2017.

[26] See the British Emergentists, Newman 1996, Bedau 1997, Silberstein and McGeever 1999, Mitchell 2012.

[27] See Morgan 1923, Sperry 1986, Klee 1984, Thompson and Varela 2001, Searle 1992, Schroder 1998, Stephan 2002.

[28] See Putnam 1967, Fodor 1974, Boyd 1980, Klee 1984, Lepore and Loewer 1989, Wimsatt 1996, Antony and Levine 1997, Batterman 1998, Aizawa and Gillett 2009, Morrison 2012.

[29] See Yablo 1992, Lepore and Loewer 1987 and 1989, Bennett 2003, List and Menzies 2009.

[30] See Batterman 1998, Wilson 2010*b*, Lamb 2015.

[31] See Morrison 2012.

[32] See Wilson 1999.

[33] See Shoemaker 2000/2001, Clapp 2001, Rueger and McGivern 2010.

[34] See Broad 1925, Hempel and Oppenheim 1948, Klee 1984, Lepore and Loewer 1989.

[35] See Newman 1996, Bedau 1997.

[36] See Chalmers 1996, Van Gulick 2001.

[37] See Batterman 2002.

of options, it's no surprise that many discussions of specifically metaphysical emergence aim primarily to taxonomize its varieties.[38]

Now, as I have previously observed (Wilson 2015*b*), though in general a thousand flowers may fruitfully bloom, this much diversity is unhelpful as regards answering the first key question, concerning the nature and varieties of specifically metaphysical emergence. It would be one thing if different accounts of metaphysical emergence targeted different phenomena. But different accounts often target the same phenomena, while disagreeing about whether these are metaphysically emergent. And when accounts do agree that a given phenomenon is emergent, there is often no clear basis for the agreement—or, relatedly, for determining whether the emergence at issue is compatible with *physicalism*, the view that all broadly scientific goings-on are, to speak schematically, 'constituted by', 'grounded in', or 'completely metaphysically dependent on' physical goings-on. More generally, the extent of variability in both content and application here might well lead one to suppose that accounts of specifically metaphysical emergence, like accounts of emergence generally, have (beyond the comparatively abstract core components) nothing systematic in common.

1.2.2 Is there actually any metaphysical emergence?

The answer to the second key question, of whether there actually is any emergence of a metaphysical variety, has also remained unclear, in large part owing to still-live concerns about whether the appearances of such emergence (or of associated levels of natural reality) are genuine. Among these concerns are that metaphysical emergence is naturalistically unacceptable; that considerations of parsimony push against taking the appearances of metaphysical emergence ontologically seriously; that the notion of metaphysical emergence is either trivially fulfilled or trivially never fulfilled; and that metaphysically emergent entities or features, were they to exist, would give rise to problematic causal overdetermination of effects already produced by dependence base entities or features.

Here the diversity of accounts of emergence again muddies the waters; for while some accounts have resources to respond to some of these concerns, the absence of any systematic treatment of the notion of metaphysical emergence renders it unclear whether the notion can survive all the various attacks. And to the extent that the in-principle viability of metaphysical emergence remains unclear,

[38] See Klee 1984, Van Gulick 2001, Stephan 2002, Gillett 2002*b*, O'Connor and Wong 2015. See also the introduction to Gibb *et al.* 2018 for discussion of the continuing diversity of approaches to emergence.

the further project of determining whether there actually is any such emergence cannot even get off the ground.

1.2.3 My aim: to provide clear, compelling, systematic answers to the key questions

The point and purpose of this book is to provide clear, compelling, and systematic answers to the two key questions of what, more precisely, metaphysical emergence is, and whether there actually is any such emergence.

In response to the first key question, I will argue that for the sort of target cases motivating the notion of metaphysical emergence, there are two and only two schemas for—schematic characterizations of—metaphysical emergence, one of which is compatible with physicalism (on the assumption that the dependence base goings-on are physical), and the other of which is not so compatible (on that assumption). And I will show that a wide range of existing accounts of metaphysical emergence plausibly aim to instantiate one or the other schema, such that much of the apparent diversity of these accounts is superficial.

In response to the second key question, I will first argue that each of these two forms of metaphysical emergence is viable—coherent, metaphysically substantive, naturalistically acceptable, such as to avoid problematic causal overdetermination, and more generally such as to vindicate and illuminate the prima facie scientific and ordinary experiential motivations for thinking that there is metaphysical emergence. I will go on to consider, for a variety of interesting actual phenomena, whether these phenomena are metaphysically emergent in one or the other of these two ways; I will argue that one form of metaphysical emergence (the sort compatible with physicalism) is actually quite common, and the other (the sort incompatible with physicalism) remains, for some cases, an open and in-principle empirically verifiable possibility, and in one special case—arguably the most important case, for creatures like ourselves—is plausibly actually instantiated.

1.3 Outline of the book

The plan for carrying out this project is as follows.

In Chapter 2 ('Two schemas for metaphysical emergence'), I present what is seen by many as the most pressing challenge to taking the appearances of metaphysical emergence as genuine—namely, the problem of higher-level causation, made salient by Jaegwon Kim in his 1989, 1993*a*, 1998, and elsewhere. Kim's general

concern is that any purported effects of higher-level features are already produced by the lower-level features upon which they minimally nomologically supervene, such that the metaphysical emergentist is committed to such effects' being problematically causally overdetermined—that is, problematically caused twice over. I argue, following discussions in Wilson 1999, 2001, 2011*b*, and elsewhere, that there are two and only two strategies of response to this problem that make sense of seemingly higher-level entities and features' being metaphysically emergent—that is, as being cotemporally materially dependent on yet also ontologically and causally autonomous with respect to dependence base (in particular, physical) entities and features. One strategy provides a schematic basis for 'Weak' (physically acceptable) emergence; the other provides a schematic basis for 'Strong' (physically unacceptable) emergence. And for each of these strategies and associated schemas, I show that a representative range of seemingly diverse accounts of emergence are plausibly seen as satisfying the conditions in one or the other schema, and thus are more unified than they appear.

Since the schemas play a large and structuring role in this book, it is worth prefiguring their content and the associated strategies for avoiding problematic causal overdetermination. On the view to be advanced and defended in what follows, a Strongly emergent feature has, on a given occasion, at least one token power not had by the base feature upon which it cotemporally materially depends on that occasion; overdetermination is avoided by denying that the base feature produces the effect (or, more weakly, by denying that the base feature produces the effect in the same way as the higher-level feature). Strong emergence is of the anti-mechanistic or anti-physicalist variety associated with British Emergentism, according to which, at certain levels of compositional complexity, fundamentally novel features and associated powers or laws come to exist (be instantiated, obtain). By way of contrast, a Weakly emergent feature has, on a given occasion, a proper subset of the token powers had by the base feature upon which it cotemporally materially depends on that occasion; problematic overdetermination is thus avoided insofar as every token power of the higher-level feature is identical to a token power of its base feature, while the distinctive efficacy of the higher-level feature is preserved as a result of its having a distinctive power profile. Weak emergence is the sort associated with nonreductive physicalism, according to which some higher-level features are, while completely metaphysically dependent on complex configurations of ultimately physical goings-on, nonetheless distinct from and distinctively efficacious with respect to the latter. For purposes of appreciating the generality of the schemes, it is worth registering that the notion of 'power' here is metaphysically highly neutral, reflecting commitment just to the plausible thesis that the causes an entity may potentially bring about are associated (perhaps only contingently) with how the entity is—that is, with its features. More

generally, no controversial theses pertaining to the nature of powers, properties, causation, or laws are presupposed.[39]

The results of Ch. 2 establish that we have prima facie reason to think that satisfaction of the conditions in the schemas for Weak and Strong emergence is, as I put it, "core and crucial" to metaphysical emergence of physically acceptable and physically unacceptable varieties, respectively. I prefer this terminology to the usual though to my mind overly coarse-grained terms of necessary and sufficient conditions, since any schematic account needs to be sensibly filled in. But modulo this caveat, the results of this chapter can also be seen as providing prima facie reason to think that the conditions in the schemas are, when sensibly filled in, both necessary and sufficient for metaphysical emergence of both physically acceptable and physically unacceptable varieties—a bold claim, but one that, as I argue in ensuing chapters, is surprisingly robust.

In Chapter 3 ('The viability of Weak emergence') I consider and respond to a representative range of objections to the schema for Weak emergence and the associated 'proper subset of powers' approach to realization presented and developed in Wilson 1999, 2009, 2010*b*, 2011*b*, and 2015*b*, among other venues.[40] These objections include that satisfaction of the conditions in the schema is compatible with anti-realism or reductionism about the purportedly emergent features, is compatible with the physical unacceptability of the emergent features, and is not necessary for physically acceptable emergence. These diverse challenges can, I argue, be answered. As we'll see, each challenge admits of one or more responses that are generally available on any sensible implementation of the schema for Weak emergence. Upon occasion, additional responses draw on features of my preferred accounts of Weak emergence—one appealing to the determinable/determinate relation (as per my 1999 and 2009, developing the proposals in Macdonald and Macdonald 1986 and Yablo 1992), according to which Weakly emergent features are determinables of lower-level realizers, and another appealing to an account of Weak emergence as involving an elimination in degrees of freedom (as per my 2010*b*, developing the proposal in Batterman 1998 and elsewhere), according to which (roughly speakly) at least one state of a Weakly emergent entity can be specified using strictly fewer degrees of freedom (independent parameters needed to specify states relevant to an entity's law-governed properties and behaviors) than are needed to specify the corresponding state of the system of entities upon which it cotemporally materially depends.

[39] As I argue in Wilson 2015*b*, and as I'll discuss further in §1.4.4, even a contingentist categoricalist Humean—that is, someone who thinks that causation is a matter of contingent regularities, and that powers are reducible to categorical, non-dispositional features—can accept powers in the metaphysically neutral sense at issue in the schemas.

[40] As will become clear, the approach to Weak emergence/realization that I endorse bears certain similarities to, but has certain crucial advantages over, the accounts of realization endorsed in Shoemaker 2000/2001 and 2007 and Clapp 2001.

In Chapter 4 ('The viability of Strong emergence'), I consider and respond to a representative range of objections to the schema for Strong emergence and the associated 'new power' approach to physically unacceptable emergence presented and developed in Wilson 1999, 2002*a*, and 2015*b*, among other venues.[41] These objections include that satisfaction of the conditions in the schema renders Strong emergence naturalistically unacceptable or 'scientifically irrelevant', is compatible with physicalism, is impossible owing to the base feature's inheriting any purportedly novel power, and is not necessary for physically unacceptable emergence. Here again, I argue that the diverse challenges can be answered. And here again, for each challenge one or more strategies of response are available on any sensible implementation of the schema for Strong emergence. Upon occasion, however, additional responses draw on features of my preferred 'fundamental interaction-relative' account of Strong emergence (as per my 2002*a*), according to which a Strongly emergent entity (feature) has at least one power that is grounded, at least in part, in a novel (nonphysical) fundamental interaction.

Having established the in-principle viability of both Weak and Strong conceptions of metaphysical emergence, I go on to put this result to work, in considering whether complex systems, ordinary (inanimate) objects, consciousness (characteristic of persons), and free will (characteristic of agents), are plausibly seen as actually either Weakly or Strongly emergent.

In Chapter 5 ('Complex systems'), I first discuss how the historical assumption that nonlinearity is a marker of fundamental novelty of the sort at issue in Strong emergence was undermined by the discovery of complex nonlinear systems which were clearly physically acceptable. I then suggest (drawing on Wilson 2002*a*) an alternative empirical criterion of Strong emergence, as located in an apparent violation of a conservation law and associated new fundamental interaction. By lights of this criterion, the Strong emergence of complex systems remains an empirically open but currently unmotivated possibility. Turning next to Weak emergence, I argue (drawing on Wilson 2010*b* and 2013*b*) that while bare appeals to common features of complex systems such as algorithmic incompressibility (Bedau 1997 and 2008), dynamic self-organization (Mitchell 2012), and universality (Batterman 1998) do not themselves provide a decisive basis for taking complex systems to be Weakly emergent, cases can be made that these or related features entail satisfaction of the conditions in the schema. Most promisingly, I argue, complex systems exhibiting universality of the sort Batterman focuses on also have (as he observes) degrees of freedom (DOF) that are eliminated relative to the systems of their composing lower-level entities, and so are Weakly emergent by lights of a DOF-based account; and I offer reason to think that certain other

[41] Variations on this approach to Strong emergence can be found in, e.g., Broad 1925, McLaughlin 1992, and O'Connor 1994.

complex systems (Bedau's gliders in the game of Life; Mitchell's flocks of birds) may also be seen as Weakly emergent by these lights. Here also I address the concern, suggested by discussions in Morrison 2012 and Lamb 2015, that complex systems involve not *fewer* but *more* DOF than their dependence base systems, associated with 'order parameters' that arise near critical points.

In Chapter 6 ('Ordinary objects'), I consider whether ordinary (inanimate) objects, of either natural or artifactual varieties, are either Weakly or Strongly emergent. I offer three motivations for thinking that such objects are 'at least' Weakly emergent in having at least one feature satisfying conditions in the associated schema: first, I argue that insofar as quantum DOF are eliminated from the specification of ordinary objects of the sort appropriately treated by classical (or 'Newtonian') mechanics, such objects are Weakly emergent by lights of a DOF-based account; second, I argue that a common conception of artifacts as associated with sortal properties and distinctive functional roles supports thinking of these as being at least Weakly emergent by lights of a functional realization account; third, I argue that ordinary objects typically have metaphysically indeterminate boundaries, which when coupled with an attractive determinable-based account of such indeterminacy, indicates that such objects are at least Weakly emergent, by lights of a determinable-based account of such emergence. While the Strong emergence of ordinary objects remains an open empirical (if not commonly endorsed) possibility, the best such case involves artifacts. More specifically, I suggest that artifacts might be Strongly emergent—if the states of consciousness that determine what powers are possessed by artifacts are themselves Strongly emergent, as is explored in Ch. 7. I close by noting that the previous results undercut the meta-ontological view, endorsed in Thomasson 2010 and elsewhere, according to which investigations into the status of ordinary objects should proceed differently from investigations into the status of special-science entities.

In Chapter 7 ('Consciousness') I turn to considering whether consciousness of the sort that we and other creatures enjoy is plausibly seen as either Weakly or Strongly emergent. Existing arguments for the Strong emergence of consciousness rely, one way or another, on the supposition that certain of the characteristic features of consciousness—notably, its subjective or qualitative aspects—lie beyond the explanatory reach of any lower-level physical goings-on. Though, as will have been established by this point, the presence of even an insuperable explanatory gap is not in itself a sufficient indicator of Strong emergence, the proponents of explanatory gap arguments take such gaps to be metaphysically significant, in reflecting not just broadly mathematical barriers to explanation such as nonlinearity, but rather that the features at issue depart so greatly from physical features that this divergence provides reasonable grounds for thinking that no physicalist account of consciousness of either reductive or nonreductive (i.e., Weakly emergent) varieties could possibly be correct. I consider the two

most promising forms of explanatory gap argument: knowledge arguments of the sort advanced in Nagel 1974 and Jackson 1982 and 1986, and the conceivability argument advanced in Chalmers 1996, 2009, and elsewhere, and I argue that each form of explanatory gap argument is uncompelling, for reasons not much previously explored. The upshot will be that while it remains an open empirical possibility that consciousness is Strongly emergent, at present we have no compelling philosophical or empirical motivation for taking this to actually be so. I go on to argue that, on the supposition that consciousness is not Strongly emergent, attention to the irreducibly determinable nature of qualitative conscious states provides good reason to see certain such states as realized in determinable-based fashion by lower-level physical states, and hence as Weakly emergent.

In Chapter 8 ('Free will'), I consider whether free will of the sort that we appear to have and to exercise is either Weakly or Strongly emergent. I start by drawing on Bernstein and Wilson 2016 to present a framework for connecting existing positions on free will—most importantly, compatibilism and libertarianism—to existing positions in the mental (more generally: higher-level) causation debates. In our paper, Bernstein and I argue that a representative range of compatibilist accounts are appropriately seen as implementing a 'proper subset' strategy relevantly similar to that implemented by nonreductive physicalists; here I extend this result to establish that the compatibilist strategy also entails satisfaction of the conditions in the schema for Weak emergence. I then argue that a representative range of libertarian accounts are appropriately seen as implementing a 'new power' strategy entailing satisfaction of the conditions in the schema for Strong emergence. This setup established, I first argue that free will of the compatibilist/Weakly emergent variety is plausibly widespread, then present a novel argument for the conclusion that at least some instances of seemingly free choice are properly taken to be of the libertarian/Strongly emergent variety. At the end of the day—and contra discussions which have primarily focused on qualitative aspects of conscious experience (as in Chalmers 1996)—I suggest that libertarian free will provides the best case for there actually being Strong emergence.

I finish up, in Chapter 9 ('Closing remarks') by calling attention to some phenomena whose status as metaphysically emergent deserves further attention, and by making some methodological observations that point towards other ways in which the present project might be profitably extended.

1.4 Operative notions

In the remainder of this chapter, I say a bit more about certain operative notions playing a regular background role in the investigations to follow. These are, first, the physical; second, levels and their individuation; third, the fundamental; fourth,

causes and powers; fifth, methodology. Those interested in cutting to the chase of later chapters are invited to use what follows as a reference section, if and when needed.

1.4.1 The physical

As above, debates over the status of some goings-on as metaphysically emergent typically take physical monism for granted, and more generally suppose that the entities and features upon which emergent goings-on depend are ultimately physical. But what is it for some goings-on to be physical?

The usual understanding is one according to which the physical goings-on are those treated by physics. Such a view reflects a transition from an a priori to an a posteriori characterization of the compositionally basic goings-on. As Crane and Mellor (1990) tell the story, the compositionally basic entities were historically characterized as having certain features definitive of matter—being impenetrable, being conserved, being such as (only) to deterministically interact, and so on. But contemporary physics has shown that the compositionally basic entities actually have few, if any, of these characteristics; hence the characterization of such entities—deemed 'physical'—is now taken to be determined a posteriori by physics. (More precisely, in the first instance the usual understanding is one according to which physics determines the extension of the 'basic' or 'narrowly' physical entities. It remains to say how the range of basic physical goings-on should be expanded so as to constitute the physical level—a task I address in §1.4.2.)

A physics-based approach to the physical, though common, faces the concern, articulated in 'Hempel's Dilemma' (acknowledging Hempel 1979), that neither current nor future physics will do so far as characterizing the domain of the (basic) physical is concerned—at least if this domain (in combination with the operative expansive resources) is supposed to serve as a substantive dialectical basis for exploring whether some form of physicalism is correct. For if the physics at issue is current physics, then (since current physics is to some extent both inaccurate and incomplete) so will be the associated domain. But if the physics at issue is future physics, then (since we don't know what future physics will end up positing) the domain of the physical will be presently indeterminate. Indeed, for all we now know, future physics might end up positing basic entities or features that are intuitively physically unacceptable—most pressingly, in being fundamentally mental, at odds with the intended conception of the physical as contrasting with panpsychism or other doctrines that neither physicalists nor their Strong emergentist rivals would accept (Papineau 1993, Loewer 2001).

Hempel's dilemma can be avoided, however (see Wilson 2006), in a way consonant with historic and contemporary debate on the status of physicalism and competing doctrines.[42] To start with the concern that future physics might end up positing fundamentally mental goings-on: this difficulty may be avoided by noting that Crane and Mellor's genealogy omits a crucial fact—namely, that physicalists have not handed over all authority to physics to determine, a posteriori, what is physical. Reflecting the historical roots of physicalism in materialism as foundationally committed to understanding mentality as nothing over and above complex material goings on, one feature has remained definitive of the term 'physical': namely, that the compositionally basic physical entities and features are not fundamentally mental—that is, do not individually either possess or bestow mentality. Hence a physics-based account of the physical should not be understood as the view that any and all entities treated by physics—current, future, or ideal—are physical, but must rather incorporate a 'No Fundamental Mentality' constraint along these lines. Strong emergentists would agree.[43] Note that this constraint, properly understood, attaches to the basic or comparatively noncomplex entities and features which are the targeted subject matter of physics, and so does not rule out configurations of physical entities or their features (what I called the 'physically acceptable' entities and features in my 2006) from individually having or bestowing mentality.

The 'No Fundamental Mentality' constraint serves, in the main, to address Hempel-style concerns with a future physics-based account of the physical. The problem remains, however, that if the entities posited by future physics are of too different a character from those posited by present physics, the present content and applicability of such an account will be compromised—an observation which lies at the heart of efforts (as in Melnyk 1997) to make sense of a present-physics account of the physical. Though, as I argue in my (2006), present-physics accounts are not ultimately sustainable, the apt concern about indeterminacy can, I believe, be accommodated by taking the (basic) physical entities to be those that are

[42] See Stoljar 2001, Dowell 2006, and Ney 2008 for discussion of alternative conceptions of the physical and associated approaches to resolving Hempel's dilemma. A full discussion of these alternatives would take us too far afield; here I'll just register that these conceptions are for various reasons (as in, e.g., Ney's conception of physicalism as an 'attitude' as opposed to a metaphysical thesis, or Stoljar's conception of the physical as ostensively defined by reference to paradigmatic inanimate objects and their components, where nothing rules out the components' being conscious) not well-suited to characterizing the compositionally basic entities as physical in terms consonant with the presuppositions of accounts of emergence.

[43] Notwithstanding the focus in Hempel's Dilemma on formulating physicalism, the dilemma also attaches to the formulation of Strong emergentism, since this view shares with (all versions of) physicalism commitment to the compositionally basic entities' being physical. Like physicalists, Strong emergentists maintain that the compositionally basic entities neither have nor bestow mentality, as per the intended contrast of Strong emergence with views, such as pan- or proto-psychism, on which consciousness is had or possessed by individual compositionally basic entities.

approximately accurately treated by present or future (in the limit of inquiry, ideal) physics.[44]

Positing the (basic) physicality of nonfundamentally mental entities treated approximately accurately by present or future versions of physics prevents physics' present failures from immediately falsifying physicalism, while providing continuous content to the account of the physical through the needed revisions. In what follows this account of the (basic) physical will be operative, though as we'll see not much will turn on the specific details of this account. The main take-home point is that there is at least one physics-based account of the (basic) physical suitable for our dialectical purposes.

One question remains: what if there are no compositionally basic entities, as per a 'gunky' world, in which everything can be further decomposed (i.e., has further proper parts)?[45] Here I am inclined to think that even if the physical entities are further decomposable, one may treat them as (effectively) compositionally basic so long as features of the composing goings-on are either inherited or irrelevant to the influence of goings-on at the physical level or above. In such a case of infinite decomposability, the 'No Fundamental Mentality' would need to be revised, perhaps along lines of a 'No Low-Level Mentality' constraint (as per Montero 2006) or along lines of an 'Only Structural-Dynamic Mentality' constraint (as per Alter 2020). In what follows, I will assume that certain physical entities are (or can be treated as) compositionally basic, but will keep track of any concerns arising from this supposition.

1.4.2 The individuation of levels

It is natural to think of metaphysical emergence in the target cases, whether natural or artifactual, as going hand-in-hand with the suggestion that emergent entities and features are 'higher-level' with respect to the 'lower-level' goings-on upon which they depend. As Rueger and McGivern (2010) observe:

> Talk of levels or layers of reality is ubiquitous in science and in philosophy. It is widely assumed, for instance, that physical, chemical, biological, and mental phenomena can be ordered in a hierarchy of levels [...]. (379)

Still, as Wimsatt (1994) notes, "the notion of a compositional level of organization is left unanalyzed by virtually all extant analyses of inter-level reduction and emergence" (203), notwithstanding that "levels and other modes of organization

[44] See Hellman and Thompson 1975, Papineau 1993, Ravenscroft 1997, Papineau 2001, and Loewer 2001 for variations on this theme.

[45] See Zimmerman 1995 and Schaffer 2003 for discussion.

cannot be taken for granted, but demand characterization and analysis" (204). How should we understand talk of levels in what follows, and most importantly, which entities and features should be taken to exist at a given level? The question is a delicate one in various dialectical respects.

To start, at the present stage of investigation—antecedent to having considered anti-realist or reductionist reasons for rejecting realist treatments of metaphysical emergence and associated multiple levels—talk of there being 'higher' and 'lower' levels should be taken to reflect the appearances, so as not to rule 'one-level' positions out of court. As Rueger and McGivern (2010) note, in the face of certain pressures, we might well wonder whether levels have "no ontological significance and are merely different modes of representing an un-layered reality" (380).

Also important for dialectical purposes is that (as previously observed) levels (or the one level, if anti-realism or reductionism turns out to be correct) be individuated so as to include any combinations or configurations of entities and features to which the anti-realist or reductionist may reasonably appeal. Suppose, for purposes of illustration, that fundamental physics is atomic, such that the basic physical entities are atoms and the basic physical relations include spatial relations and pairwise bonding relations between atoms. Then, beyond the atoms and atomic relations, we should allow as existing, at the atomic level, not just small numbers of atoms standing in atomic relations, but also (among other aggregative combinations) large numbers of atoms standing in highly complex atomic (including spatial) relations, constituting pluralities or aggregates of the sort that might, if reductionism is correct, be identical with a rock, a plant, or a person, at least at any given time. Similar resources are needed to make sense of anti-realist views, such as Heil's (to be discussed further in Ch. 3), which reject both reductionism and the independent posit of higher-level features.[46]

If we are suitably generous to the anti-realist or reductionist, substantive debate can proceed over whether some apparently higher-level entity or feature really is higher-level, or is rather identical to some lower-level entity or feature (or perhaps, as on Heil's account, doesn't exist at all). Being suitably generous means that we cannot rest with certain ways of individuating levels. For example, Wimsatt's (1994, 226) understanding of "compositional levels of organization [...] as constituted by families of entities usually of comparable size and dynamical properties, which characteristically interact primarily with one another" rules reductionism out of court. Indeed, unless the compositionally basic level can contain complex configurations of (perhaps vastly) different sizes, one will not be able to make

[46] As Heil (2003*a*) puts it:

> I am inclined to think that 'this is a statue' can be, and often is, literally true. What makes it true is a complex, dynamic, arrangement of particles [...] We deploy predicates like 'is a statue' to mark off salient features of the world. These features are grounded in properties and arrangements of the fundamental constituents. (217)

sense of Wimsatt's reductionist-friendly claim that "Because any complex material objects can be described at a number of different levels of organization, identity relations must hold between descriptions of the same object at different levels" (227–8).[47] For related reasons, the occasional suggestion in the literature that it is indicative of a new 'level' (or more specifically, of emergence) that a feature of a configuration be different, one way or another, from any feature of the individual entities entering into the configuration (as when the mass of a composite has a different value from the mass of any of its parts, or when a diamond has a property, such as hardness, that no carbon atom has), is unsatisfactory for present dialectical purposes; for a reductionist can happily agree that configurations (pluralities or structural aggregates) have features not had by the individuals entering into the configurations.[48]

On the other hand, neither do we want to be so generous to the reductionist (or other deflationist) as to rule the possibility of either Weak or Strong emergence out of court.

With these constraints in mind, the question to be answered is: which combinations of entities and associated features should be taken to exist at a given level *L* of broadly scientific reality—most crucially, the fundamental physical level of reality—beyond the entities and features typically taken, by lights of the associated science *S*, to be characteristic of *L*? In short: how should levels be individuated? In what follows, I'll discuss two different approaches to answering this question, neither of which is perfect, but either of which suffices to get discussion off the ground.

The 'lightweight combination' approach

One common approach to the individuation of levels proceeds by allowing that various ontologically 'lightweight' combinations of the characteristic entities and features treated by a given science *S* and placed at a level *L* are also appropriately placed at *L*. For example, Hellman and Thompson (1975) first individuate the compositionally basic level in terms of the entities and features taken to be characteristic of fundamental physics, as including any

[47] That said, Wimsatt does aim to make sense of such identity claims by reference to a number of modes of 'aggregation', in a way that is friendly to the 'lightweight combination' approach to be discussed shortly.

[48] Discussions of emergence in scientific contexts, when not appealing merely to epistemic notions, frequently appeal to these sorts of conceptions of emergence, tracking novelty of features of a whole with respect to features of the parts (as in, e.g., the characterization in Blitz and Richmond 1994 of "emergence in a system [as involving] a property not possessed by any of its parts", 1)—though these discussions typically then walk this back by introducing aggregative considerations which more accurately capture the import of the oft-cited slogan that 'the whole is more than *the sum* of the parts'.

> … satisfying any predicate in a list of basic positive physical predicates of [fundamental physical theory]. Such a list might include, e.g., 'is a neutrino', 'is an electromagnetic field', 'is a four-dimensional manifold', and 'are related by a force obeying the equations (Einstein's, say) listed', etc. (554)

(See also Melnyk 1997 and others.) Hellman and Thompson then expand beyond the entities and features picked out by the basic physical predicates to include at the basic physical level (to simplify somewhat) any and all mereological sums (i.e., 'fusions' of parts into wholes) of spatiotemporally located instances of basic physical entities and features.[49]

Several other modes of combination are usually allowed, applying to entities, features, or both, which are supposed by all parties not to result in any interesting form of emergence. These are typically operations which are uncontroversially aggregative using resources of the science at issue (e.g., spatial relations, or iterations of pairwise relations, as in the atomic example just given), or are broadly logical, as involving certain Boolean, classical mereological, or set-theoretic combinations of entities or features. Some mathematical modes of combination are also considered ontologically lightweight in this context; in particular, it is common to suppose that features corresponding to linear combinations of L-level features should also be placed at L. Hence at a level L associated with a given science S, the L-level entities and features (including properties, states, and relations) would typically be taken to include:

- any characteristic entities or features treated by S
- any entity (or feature) consisting in a set or plurality of L-level entities (or features), understood as (merely) jointly existing
- any entity (or feature) consisting in a disjunction of any L-level entities (or features)
- any entity (or feature) consisting in a conjunction of L-level entities (or features)
- any entity (or feature) consisting in a mereological fusion of L-level entities (or features)
- any feature consisting in a linear (scalar or vector) combination of L-level features[50]

[49] Classical mereology is a theory of parts and wholes on which composition ('fusion') is ontologically lightweight, in that any objects *qua* parts automatically form a whole. There is controversy over whether material composition (as when, e.g., some atoms compose a molecule, and so on) can or should be understood in terms of mereological composition; see Simons 1987, van Inwagen 1990, McDaniel 2001, Paul 2002, Koslicki 2008, Bennett 2015, and Varzi 2019 for discussion.

[50] Wimsatt (1994) more generally suggests that "the conditions required for a system property to be an aggregate of the properties of the parts of the system—conditions on the 'composition function'

- any relational entity consisting in any number of L-level entities standing in L-level relations
- any relational feature of the form "being composed of some L-level entities with L-level features standing in L-level relations."[51]

Note that the specification here allows for iterative closure under the operations; for example, an entity consisting in a disjunction of conjunctions of L-level relational entities would also be L-level. Though in principle there might be further ontologically lightweight operations, it is common to suppose, on this approach, that the closure of L-level entities under these operations is more or less exhaustive of the (individual and aggregative) entities at L. Hence it is that debate over the status of a given seemingly higher-level entity or feature frequently proceeds by considering whether the entity or feature can be reduced to one or other of these lightweight combinations of characteristic lower-level entities or features. For example, a nonreductive physicalist might argue that the multiple realizability of a given mental feature type indicates that it is not reducible to any type of lower-level physical feature, while a reductive physicalist might respond by suggesting that the multiple realizability of the mental type can be accommodated, compatible with reduction, by taking the mental type to be identical to a disjunction of physical types.

For most purposes in what follows, no harm comes from taking levels to be individuated along lines of the ontologically lightweight approach, as again is commonly done; I will flag any exceptions to this rule.

relating system and parts' properties" are "associativity, commutativity, inter-substitutivity, linearity, and invariance under decomposition and reaggregation" (237).

[51] Features of this form are sometimes called "micro-based" or "micro-structural" features; see Kim 1998 and Shoemaker 2007, following Armstrong 1978. Kim (1998) offers on the reductionist's behalf the feature of being a water molecule as a case in point: "it is the property of having two hydrogen atoms and one oxygen in such and such bonding relationship" (84). More generally, Kim (1998) characterizes a micro-based property as follows:

> P is a micro-based property just in case P is the property of being completely decomposable into nonoverlapping proper parts, a_1, $a_2, \ldots, a_n$, such that $P_1(a_1)$, $P_2(a_2), \ldots, P_n(a_n)$, and $R(a_1, a_2,, \ldots, a_n)$. (84)

Shoemaker's related (2007) characterization is as follows:

> [Micro-structural] properties [...] can be specified entirely in terms of the micro-manifest powers of the constituent micro-entities together with how these micro-entities are related i.e., in terms of what could be known about them prior to their entering into emergence engendering combinations. Such a property will be the property of being composed of particles with such and such micro-manifest causal powers and related in such and such a way. [...] If emergentism is false, manifest causal powers are the only ones the micro-entities have, and physical micro-structural properties are the only ones macro-objects have, and the other properties of macro-objects are realized in their physical micro-structural properties. (55)

The law-consequence approach

An alternative approach to the individuation of levels, which I here put on the table, expands upon the appeal to scientific laws, understood as applying to entities at a distinctive level of natural reality. Laws so understood are, in the first instance, metaphysical; they are, or encode, the 'rules' governing the entities and features at issue, though, consonant with a suitably fallibilist realism about theories, discussion of laws typically focuses on claims made in or by scientific theories as appropriate stand-ins for claims about the associated laws.

On the law-consequence approach, the suggestion is that the laws governing entities characteristic of a given level can also do the work of expanding the domain of entities and features at that level in such a way that an anti-realist or reductionist view is not ruled out of court. The laws of fundamental physics, for example, are capable of taking as input or initial conditions various complex configurations (pluralities or structural aggregates) of characteristic physical entities and features; hence the laws/theories themselves have resources to expand beyond the explicit focus of the laws/theory treating some *L*-level entities and features, to admit at that level any entities and features whose (potential) existence is deemed a metaphysical consequence—not to be confused with either mere necessitation or representational entailment—of the *L*-level laws.

A law-consequence approach to the individuation of levels has certain advantages over a lightweight combination approach. For example, a law-consequence approach need not antecedently specify whether nonlinear entities or features are or are not to be placed at a given level *L*; whether this is so will follow from the laws governing the characteristic entities at that level. Relatedly and more generally, on this approach entities and features which are causal consequences just of the *L*-level laws may also be placed at *L*. Another advantage of a law-consequence approach is that it need not be committed to any and every ontologically lightweight combination of entities or features at *L*'s also being placed at *L*—which exclusion might be desirable if, say, some complex configurations of atoms could not exist, for some law-based reason.

That said, one might be concerned that, while a law-consequence approach to levels clearly makes room for the possibility of Strong (physically unacceptable) emergence (since fundamental higher-level powers or other features of reality will not be metaphysical consequences of lower-level laws), it rules out Weak (physically unacceptable) emergence as associated with nonreductive physicalism. After all, nonreductive physicalists grant that the higher-level entities and features that they take to be genuine, as well as the higher-level laws governing these entities and features, are *in some sense* metaphysical consequences of the fundamental physical laws, even if these higher-level goings-on are (as nonreductive physicalists suppose) different from any lower-level goings-on (and moreover, as some nonreductive physicalists think, are in some cases epistemically beyond our ken).

However, a law-consequence approach, properly understood, does not rule out (the possibility of) Weak emergence. For if there are any metaphysical consequences of the sort that the Weak emergentist supposes there to be, these will not be appropriately placed at *L*. In Ch. 3, I'll develop this point in terms of degrees of freedom. The basic idea is intuitive, however. To start, laws require certain kinds of information in order to operate. For example, in order for the quantum laws to operate on a configuration of fundamental physical particles, the input into the laws has to contain information about the individual quantum spins of those particles. Suppose also that it is a metaphysical consequence of the quantum configuration's existing and evolving as per the quantum laws that a given planet exists and evolves as per the laws of astronomy, where (as a Weak emergentist might well maintain) the characterization of planets and their behaviors is taken to abstract away from any details concerning quantum spin. Were the planet, so characterized, to be given as input into the quantum laws, this input would not contain the information about quantum spin required for the laws to operate. Correspondingly, and more generally, planets and their features would not be appropriately placed at the fundamental physical level, since the physical laws would not be able to operate on such entities or features.

The upshot is that a law-consequence approach to the individuation of levels does not entail that any and all metaphysical consequences of the laws at a given level *L* should be placed at that level. Rather, only those consequences which preserve the degrees of freedom required to allow the laws at level *L* to operate should be placed at that level. Nor does a law-consequence approach to the individuation of levels build in the truth of emergentism: whether there are actually any goings-on whose characterization eliminates lower-level degrees of freedom is a matter for empirical and philosophical determination, to be further considered down the line.

1.4.3 The fundamental

In what follows, we will have occasion to consider some goings-on—entities, features, interactions—as fundamental. For example, both physicalists and their Strong emergentist rivals suppose that the physical goings-on are fundamental, at least to the extent that certain physical entities are compositionally basic; where they disagree is over whether there are any *other* fundamental goings-on. But what is it for some goings-on to be fundamental?

There are three main approaches to this question (see Tahko 2018 for discussion), divided into primitivist and non-primitivist varieties. Non-primitivist approaches aim to metaphysically analyze what it is for some goings-on to be fundamental in terms of dependence, one way or another. On independence-based

conceptions, what makes it the case that some goings-on at a world *w* are fundamental at *w* is that those goings-on are (individually) metaphysically independent at *w*.[52] On dependence-based conceptions, what makes it the case that some goings-on at a world *w* are fundamental at *w* is that those goings-on are part of a complete minimal basis for everything that exists at *w*, such that everything that exists at *w* is either fundamental or completely metaphysically dependent on what is fundamental.[53] By way of contrast, on primitivist accounts, what makes it the case that some goings-on at a world *w* are fundamental at *w* is not metaphysically analyzable in other terms, positive or negative.[54]

My own preference is for a primitivist account. I follow Fine (2001) in thinking that one should "reject the idea that the absolute notion of fundamental reality is in need of a relational underpinning" (25); as I previously put it, the fundamental goings-on "play a role analogous to axioms in a theory—they are basic, they are 'all God had to do, or create' " (Wilson 2014*a*). (Nota bene that it is not the fundamenta themselves, but *what makes it the case that some goings-on are fundamental*, that is taken to be primitive.) Such an account has (so I argue in Wilson in progress*b*) numerous advantages over either variety of non-primitivist account. For the most part, however, this dispute will not matter for what follows.

The main exception to this rule concerns the usual understanding of an independence-based conception, according to which a given entity *x* is fundamental just in case it does not depend on any other entity (see, e.g., Schaffer 2009, 373 and Bennett 2017, 138). Such a conception problematically rules out numerous accounts of fundamental structure, including

- Cases involving self-dependent phenomena (as per a self-sufficient God)
- Cases involving fundamenta which metaphysically depend on some but not all other fundamenta (say, quarks, which in coming only in pairs or triplets appear to be existentially dependent on other quark types and/or tokens)[55]
- Cases involving fundamenta each of which metaphysically depends on all other fundamenta (as per Hua-yen Buddhism or ontic structural realism)

and most importantly for present purposes,

- Cases involving fundamenta that are partly but not completely metaphysically dependent on other fundamenta (as per typical understandings of Strongly emergent phenomena).

[52] See, e.g., Schaffer 2009, Audi 2012*a*, Dasgupta 2014, Raven 2016, and Bennett 2017.
[53] See, e.g., Sider 2011, Jenkins 2013, Paul 2002, and Tahko 2014.
[54] See, e.g., Fine 2001 and Wilson 2014*a* and in progress*b*.
[55] See Tahko 2018 and Wilson 2020 for discussion.

Whether some other version of an independence-based conception of fundamentality might make room for such Strong emergence remains to be seen. Here it is worth noting, however, that proponents of independence-based conceptions do not distinguish between the claim that *individual* fundamenta do not depend on any other goings-on, and the claim that the fundamental goings-on *collectively* do not depend on any other goings-on. The former claim seems to me to be unmotivated, given the many seeming counterexamples, whereas the latter claim seems true and capable of accommodating the typical understanding of Strongly emergent entities or features. Correspondingly, appeals to fundamentality in what follows, as applying either to physical or to Strongly emergent phenomena, may be understood as broadly neutral between a primitivist, a dependence-based ('complete minimal basis'), or a collective-independence-based conception of fundamentality.

1.4.4 Causes and powers

The discussions to come often advert to causal relations and associated powers to produce effects, as had by both physical and macro-level (special-scientific or artifactual) goings-on. There are vast literatures on causation and powers, as well as on how these notions enter, metaphysically and modally, into the characterizations of entities and features. Fortunately, for purposes of the present project it is possible to remain almost entirely neutral as regards these more specific details.

To start, the operative notion of 'power' in what follows is metaphysically highly neutral, following the presuppositions operative in my (2015*b*):

> [T]alk of powers is simply shorthand for talk of what causal contributions possession of a given feature makes (or can make, relative to the same laws of nature) to an entity's bringing about an effect, when in certain circumstances. That features are associated with actual or potential causal contributions ('powers') reflects the uncontroversial fact that what entities do (can do, relative to the same laws of nature) depends on how they are (what features they have). So, for example, a magnet attracts nearby pins in virtue of being magnetic, not massy; a magnet falls to the ground when dropped in virtue of being massy, not magnetic. Moreover, a feature may contribute to diverse effects, given diverse circumstances of its occurrence (which circumstances may be internal or external to the entity possessing the feature). Anyone accepting that what effects a particular causes (can cause, relative to the same laws of nature) is in part a function of what features it has—effectively, all participants to the present debate—is in position to accept powers, in this shorthand, metaphysically neutral and nomologically motivated sense. (354)

Note that whether internal or external circumstances are at issue, the sense in which a given feature (potentially) contributes to causing a given effect in certain circumstances is minimally one according to which those circumstances alone are not up to the task of (potentially) causing that effect.

Relatedly, the operative notion of causation in what follows is also metaphysically highly neutral. Again, there are many specific accounts of this notion, but it may serve as an initial proof of concept (to be filled in a bit in Ch. 2, after presenting the two schemas for emergence) that even a contingentist categoricalist Humean—someone who thinks that causation is a matter of regularities, such that features have their powers contingently, and that all features are ultimately categorical—can accept powers and the associated notion of causation in the neutral sense(s) here: for such a Humean, to say that an (ultimately categorical) feature has a certain power would be to say that, were a token of the feature to occur in certain circumstances, a certain (contingent) regularity would be instanced. Contemporary Humeans implement more sophisticated variations on this theme; but the point remains that no 'heavyweight' notion of powers or causation need be presupposed in what follows.[56]

More generally, no controversial theses pertaining to the nature of powers, causation, properties, or laws are here presupposed. Four points of non-commitment are worth highlighting. First, nothing in what follows requires accepting that it is essential to features that they have the powers they actually have—that is, that they make the causal contributions they actually do, when in certain circumstances. Perhaps features are essentially (or necessarily) associated with certain powers; perhaps they aren't. (I do assume, as is uniformly done by both 'contingentists' and 'necessitarians', that features have the same powers relative to the same laws of nature.) Second, nothing in what follows requires accepting that features are exhaustively individuated by powers. Perhaps they are, perhaps they aren't; perhaps features are also or ultimately individuated by quiddities or other noncausal aspects of features. Third, nothing in what follows requires accepting that powers are or are not reducible to categorical (non-dispositional) features, or that attributions of powers are or are not reducible to certain conditionals or counterfactuals, etc. Perhaps powers, or talk of them, are reducible to other entities or terms; perhaps they aren't. Fourth, nothing in what follows requires endorsing a specific understanding of features—as, e.g., universals (of some variety), or rather tropes (of some variety). And as per usual, nominalists rejecting properties or

[56] Carruth (2018) argues that some heavyweight conceptions of powers might not be suited for certain emergentist purposes, since the conceptions open the door to certain reductionist strategies. A full engagement with Carruth's discussion would take us too far afield; here I'll just register that certain of the reductionist strategies he advances fail in the face of methodological considerations to be discussed shortly, and others are along lines of those operative the 'collapse objection' to be treated in Ch. 4 (§4.2).

other features are invited to implement their preferred metaphysical translation strategy.

One last issue pertaining to the operative understanding of causation and powers deserves mention here. Both Kim's problem of higher-level causation and the schemas for metaphysical emergence motivated by this problem presuppose that there is physical causation and associated powers—but some (notably, Russell, 1912 and Field, 2003) maintain that such a supposition is problematic. As Field characterizes Russell's line of thought, "not only does the word 'cause' not appear in the advanced sciences, but the laws that these sciences state are incompatible with causation as we normally understand it" (435). To be sure, that there is no physical causation remains an outside view, as is reflected in the standard understanding of physicalism as committed to *Physical Causal Closure* (to be discussed in more detail in the next chapter), according to which any physical effect has a sufficient purely physical cause; so I could perhaps be forgiven for simply bracketing the concern. But it may be worth briefly registering why I take the assumption of physical causes and powers to be justified.

To start, that physical equations are expressed in broadly mathematical rather than explicitly causal terms seems to me to be best understood as reflecting the usual scientific focus on the prediction of quantities, coupled with a kind of expressive underdetermination associated with logical and mathematical expressions, as opposed to encoding that physical goings-on do not enter into causal relations. In '$F = ma$', to take a toy example, is the import of the '=' sign causal or not? What matters for this and related questions is whether the usual descriptions and interpretations of physical theories make reference to causal notions; and on the face of it (as in the quotation by Kane near the start of this chapter) they frequently do so, in adverting to interactions, transfers of energy, behaviors, and other notions naturally seen as having causal connotations (see Earman 1976 for similar observations).

A more serious concern would be if the laws of physics were "incompatible with causation as we normally understand it". But here too doubts can be raised. To start, Russell's argument for this conclusion presupposes a certain implausible conception of causation, according to which true causal claims take the form of universal generalizations. Here the fact that the notions of causation and powers operative in what follows are metaphysically highly neutral is relevant, since for all Russell establishes there may be alternative conceptions of physical causation compatible with our best science and one's preferred metaphysical commitments. Field discusses another potential source of incompatibility, according to which (unlike physical laws) causation is 'directed', such that causes typically precede and determine their effects, rather than *vice versa*; but as he notes, there are available strategies for locating directionality in the physical laws.

Field ends up taking the main concern about physical causation to be that "since there is always a possibility of interventions from afar", a specification of the cause of an effect up to the task of guaranteeing the associated effect will have to include "each part of the past light cone of an event among the causes of the event" (439). But this expression of the concern reverts to Russell's implausible conception of a cause as that which guarantees the associated effect (such that causal claims take the form of universal generalizations). In fact, that causal relations are typically *ceteris paribus*—i.e., interfering factors aside—is par for the course so far as causal relations in the special sciences are concerned, and from what I can tell there is no barrier to understanding physical causation as similarly admitting of exceptions. I conclude that it is the supposition that physical causation would have to be exceptionless, rather than the notion of physical causation itself, that is "a relic of a bygone age".

1.4.5 Methodology

Notwithstanding the Cartesian caricature of metaphysicians as aiming to discover the nature of reality from the comfort of their armchairs, contemporary metaphysicians more typically endorse a broadly abductive methodology, also known as 'inference to the best explanation' (see, e.g., Harman 1965 and Douven 2017), whereby candidate metaphysical accounts of a given phenomenon are assessed by attention to how well they do, overall, at satisfying various theoretical desiderata. As Sider (2009) describes 'mainstream metaphysics':

> Competing positions are treated as tentative hypotheses about the world, and are assessed by a loose battery of criteria for theory choice. Match with ordinary usage and belief sometimes plays a role in this assessment [...] Theoretical insight, considerations of simplicity, integration with other domains (for instance science, logic, and philosophy of language), and so on, play important roles. (358)

Notwithstanding this broadly common metaphysical methodology, there is considerable variation in exactly which theoretical desiderata are operative (i.e., how to fill in Sider's reference to 'and so on') as well as in how these desiderata, which may push in different directions, should be weighted.

Though some see the lack of current consensus as regards fine-grained metaphysical methodology as reason to be somewhat skeptical about metaphysical progress, my own view is that (as discussed in Wilson 2011*a*, 2016*b*, and 2016*c*) there is at present no warrant for such pessimism. The lack of current fixed standards is unsurprising, given the wide purview of metaphysical investigations

as potentially open to any and all data relevant to a given phenomenon, and given our present distance, metaphysically and epistemologically, from the end of inquiry. There is moreover reason to think that we are slowly but surely making methodological advances. Indeed, the present common acceptance of inference to the best explanation as operative in philosophy as well as in the sciences itself reflects an advance in methodological understanding, associated with the demise of verificationism in particular. In any case, in circumstances where there is some but not complete consensus regarding methodological standards, progress can nonetheless be made, so long as one is suitably explicit about which theoretical desiderata are primarily guiding one's investigations. If one is so explicit, the hope is that even those with somewhat different methodological sensibilities may find one's investigations useful, in mapping the domain of associated options as regards understanding the phenomenon at issue. Here, then, I put on the table two methodological desiderata which will guide the investigation to follow.

The first is that we should aim to realistically accommodate the appearances of metaphysical emergence, in the absence of specific reason(s) to think that this cannot be done. To be sure, we have yet to consider the range of objections that might be and have been raised against such emergence. Even so, in what follows I take accommodation of these appearances on their own terms—that is, accommodation of the natural (i.e., straightforward, default) understanding of the relevant scientific and ordinary experiential data, as suggesting that there are phenomena coupling cotemporal material dependence with ontological and causal autonomy—to be a heavily weighted virtue of an account of this data. As I see it, we should give up such natural understandings of the structure of reality as admitting metaphysically emergent entities and features only if forced to do so—e.g., by causal overdetermination considerations of the sort to be considered in Ch. 2, or by arguments showing that reasonable conditions on emergence cannot be satisfied. In particular, I do not take a supposed desire for ontological parsimony for its own sake, of the sort that might lead one to attempt to 'explain away' the prima facie motivations for metaphysical emergence even in the absence of specific objections to this notion, as having a similar weight. (That's not to say that parsimony considerations don't carry any weight, of course.) The intended desideratum can then be expressed as follows:

> *Criterion of Appropriate Accommodation*: An adequate account of metaphysical emergence should make natural (straightforward, default) and realistic sense of the appearances of metaphysical emergence, in the absence of specific reasons to think that this cannot be done.

Reflecting the *Criterion of Appropriate Accommodation*, though I do take it to be a central part of the burden in what follows to show that various purported problems with metaphysical emergence can be addressed, I do not take it to be part of the burden to show that *no* deflationary (anti-realist or reductionist) account of the appearances of metaphysical emergence (or of the considerations motivating these appearances, including, e.g., there being seemingly distinctive special-science taxonomies and laws) is viable. My ultimate goal, in other words, is not to knock the anti-realist or reductionist off their horse, but rather to show the metaphysical emergentist who aims to accommodate the appearances at realistic face value how to stay on their own horse. Again, I hope that those with different methodological sensibilities will nonetheless find the ensuing discussion useful, at least as revealing the extent to which the antecedent weighting of parsimony considerations, as opposed to any specific problem with the notion of metaphysical emergence itself, may be playing a role in deflationary accounts of such emergence.

A second criterion reflects my supposition that (again, modulo reasons to think that no realistic account of the appearances can be given) an adequate account of metaphysical emergence must not only appropriately accommodate these appearances, but moreover do so in a way that provides an illuminating—that is, explanatorily relevant—basis for understanding just how the characteristic features of such emergence might be in place:

> *Criterion of Illuminating Accommodation*: An adequate account of metaphysical emergence should provide an illuminating basis for accommodating the appearances of metaphysical emergence in natural (straightforward, default) fashion.

Relatedly, an adequate account of metaphysical emergence must provide a clear basis for the in-principle resolution of disputes over whether some phenomenon is metaphysically emergent, in a way going beyond appeal to brute intuitions or irrelevant distinctions. For example (as I noted in discussing a version of this criterion in Wilson 2019), even supposing that an oracle existed who could infallibly report whether a given phenomenon is metaphysically emergent, an account pitched in terms of such oracular pronouncements would fail to illuminate what it is to be metaphysically emergent, and so would fail to satisfy the criterion. More plausibly, an account according to which metaphysically emergent phenomena are cotemporally materially dependent on but ontologically and causally autonomous with respect to lower-level physical phenomena, but which did not provide any

explanatory insight into how, exactly, such dependent goings-on might be so autonomous, would lead directly to stalemate between emergentists and their rivals (e.g., with regard to the status as emergent of certain mental states), and so would fail to satisfy the *Criterion of Illuminating Accommodation*.

Our operative notions on the table, let us begin.

2

Two schemas for metaphysical emergence

In this chapter, I start by presenting what is seen by many as the most pressing challenge to taking the appearances of emergent structure as genuine—namely, the problem of higher-level causation, made salient by Jaegwon Kim in his 1989, 1993*a*, 1998, and elsewhere, according to which metaphysically emergent higher-level features would problematically causally overdetermine effects already brought about by the base features upon which they cotemporally materially depend.[1] I argue, following discussions in Wilson 1999, 2011*b*, 2015*b*, and elsewhere, that there are two and only two strategies of response to this problem that appropriately accommodate there being metaphysically emergent entities and features—that is, entities and features that are cotemporally materially dependent on, yet also ontologically and causally autonomous with respect to, dependence base entities and features. One of these strategies provides a schematic basis for 'Weak' metaphysical emergence, of the sort that would be compatible with physicalism, on the assumption that the compositionally basic entities are physical; the other provides a schematic basis for 'Strong' metaphysical emergence, of the sort that would be incompatible with physicalism, on that assumption.[2] Along the way, I show that representative ranges of seemingly diverse accounts of realization (i.e., Weak emergence) and of Strong emergence are plausibly seen as providing an illuminating basis for satisfying the conditions in the relevant schema, such that accounts of metaphysical emergence are more unified than they appear.

I conclude that we have prima facie reason to think that satisfaction of the conditions in the schemas for Weak and Strong emergence is, as I put it, 'core and crucial' to metaphysical emergence of physically acceptable and physically unacceptable varieties, respectively. Of course, any schematic account needs to be filled in and moreover filled in sensibly, if it is to be really adequate. But modulo this caveat, the results of this chapter can also be seen as providing prima facie

[1] See Malcolm 1968 for an early variation on the theme, focused on an apparent conflict between mechanical and intentional (psychological) explanations. Recall that 'emergent' and 'higher-level' are not exact synonyms, given that vitalist and substance pluralist views of the sort intended to contrast with any form of emergentism also countenance higher-level entities and features.

[2] As previously noted, although the assumption that the base-level entities and features are physically acceptable is typically operative in what follows, the schemas generalize to characterize emergence of two different varieties, whatever the precise ontological status of the base-level goings-on.

Metaphysical Emergence. Jessica M. Wilson, Oxford University Press (2021). © Jessica M. Wilson.
DOI: 10.1093/oso/9780198823742.003.0002

reason to think that the conditions in the schemas for Weak and Strong emergence are, when sensibly filled in, both necessary and sufficient for (appropriate and illuminating accommodation of) metaphysical emergence of physically acceptable and physically unacceptable varieties, respectively—a bold claim, but one that, as we will see in ensuing chapters, is surprisingly robust.

2.1 The problem of higher-level causation

I start with three clarificatory remarks. First, following Kim and standard practice, I assume that entities (objects, systems, or other particulars) are efficacious in virtue of having efficacious features (states, properties, behaviors, or other ways for entities to be). For example, the effects that a billiard ball causes (can cause) are a matter of what features it has—its mass, shape, velocity, and so on. Correspondingly, in what follows talk of entities' causing effects is suppressed in favor of talk of their features' causing effects. The assumption that the efficacy of entities lies in their having efficacious features is conveniently consonant with the usual assumption that the emergence of entities is ultimately a matter of the emergence of certain features. Second, again following Kim and standard practice, I assume that causation is in the first instance a relation between spatiotemporally located goings-on; hence reference to 'features' in what follows is to be understood, unless otherwise qualified, as reference to spatiotemporally located tokens (e.g., particular states or property instances, or associated particular events) of a given (state, property, event) type.[3] Third, to fix ideas I set up the problem as directed at special-science entities and features, but nothing deep hangs on this focus; the problem arises more generally for higher-level entities and features of either natural or artifactual varieties.

2.1.1 Kim's overdetermination argument

Six premises lead to the problem of higher-level causation.[4] Four of these concern special-science features, and are motivated by considerations similar to those giving rise to there seeming to be metaphysical emergence. These are:

[3] That said, I will sometimes gloss the type/token distinction—e.g., when discussing the necessitation of one feature by another, later in this section.

[4] What follows reflects my preferred way of presenting the problem and slate of candidate resolutions, as set out, e.g., in Wilson 2009, 2011*b*, 2015*b*. As I'll discuss, Kim's own presentations aim to motivate reductive over nonreductive versions of physicalism.

1. *Dependence*. Special-science features cotemporally materially depend on lower-level physical features (henceforth, 'base features') in such a way that, at a minimum, the occurrence of a given special-science feature on a given occasion minimally nomologically supervenes on base features on that occasion.[5]
2. *Reality*. Both special-science features and their base features are real.
3. *Efficacy*. Special-science features are causally efficacious.
4. *Distinctness*. Special-science features are distinct from their base features.

The remaining two premises concern causation. The fifth is a standard physicalist commitment, sometimes called 'the causal closure of the physical':

5. *Physical Causal Closure*. Every lower-level physical effect has a sufficient purely lower-level physical cause.[6]

The sixth reflects the common supposition that there is no systematic causal overdetermination (henceforth, just: 'overdetermination') of effects by distinct individually sufficient causes, with the exception of cases involving causes forming part of a single diachronic causal chain, and 'double-rock-throw'-type cases, where a given effect (e.g., a window's breaking) is, on a given occasion, the result of two distinct causes (e.g., two rock-throwings), each of which is individually sufficient for an effect of the type at issue. Note that in neither type of 'exception' case does one of the competing causes stand in a relation of cotemporal material dependence to the other.

6. *Non-overdetermination*. With the exception of cases of the double-rock-throw variety, effects are not causally overdetermined by distinct individually sufficient cotemporal causes.[7]

[5] Recall: by 'minimal nomological supervenience' is meant that in worlds with relevantly similar laws of nature, any given token of the supervenient (e.g., special-scientific) type requires, for its occurrence, a token of some base (e.g., physical) type; and in such worlds, if any token of that base type occurs, then a token of the supervening type will occur.

[6] In being formulated in terms of lower-level physical causes and effects, this characterization of *Physical Causal Closure* is similar to those in Baker 1993, 79 ("Every instantiation of a micro-physical property that has a cause at t has a complete micro-physical cause at t" and Sturgeon 1998, 124 ("Every quantum event has a fully disclosive, purely quantum history"). See Montero 2003 and Garcia 2014 for discussion of more generous characterizations of the closure principle in terms of any (even higher-level) physically acceptable features, which characterizations are not useful for purposes of setting up Kim's problematic.

[7] This characterization of *Non-overdetermination* (also sometimes called 'the causal exclusion principle') is along lines of that in Yates 2012, 3 ("No single event can have more than one sufficient cause occurring at any given time—unless this is a genuine case of causal over-determination").

On to the problem. There are two cases to consider, reflecting two sorts of effect that might be at issue. In the first case, a special-science feature S is assumed to cause another special-science feature S^*; in the second case, S is assumed to cause a base feature P^*. In Kim's classic presentation, S is taken to be a mental state (e.g., a state of being thirsty); P is taken to be a base state upon which mental state S depends; and mental state S is taken to cause either another mental state S^* (e.g., a desire to quench one's thirst) or a base state P^* (e.g., a physical reaching for a glass of water). More generally, however, the considerations to follow raise a concern about how any real and distinct higher-level feature might be unproblematically efficacious.

First (case 1), suppose that S causes special-science feature S^* on a given occasion (compatible with *Efficacy*). S^* is cotemporally materially dependent on some base feature P^* (*Dependence*), such that P^* necessitates S^*, with at least nomological necessity. Moreover, P^* has a sufficient purely lower-level physical cause (*Physical Causal Closure*)—plausibly, and without loss of generality, P. If P causes P^*, and P^* (at least nomologically) necessitates S^*, then it is plausible that P causes S^*, by causing P^*. So, it appears, both P and S cause S^*, and given that P and S are both real and distinct (*Reality*, *Distinctness*), S^* is causally overdetermined; moreover (given *Dependence*), this overdetermination is not of the double-rock-throw variety (contra *Non-overdetermination*). Diagrammatically, the case is as in Figure 2.1, with bold lines representing causation:

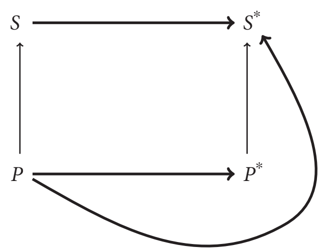

Figure 2.1 Case 1 of the problem of higher-level causation: S causes S^*

Second (case 2), suppose that S causes some base feature P^* on a given occasion (compatible with *Efficacy*). P^* has a sufficient purely lower-level physical cause (*Physical Causal Closure*)—plausibly, and without loss of generality, P. So, it appears, both P and S cause P^*, and given that P and S are both real and distinct (by *Reality* and *Distinctness*), P^* is causally overdetermined; moreover, (given *Dependence*) this overdetermination is not of the double-rock-throw variety (contra *Non-overdetermination*). Diagrammatically, the case is as in Figure 2.2:

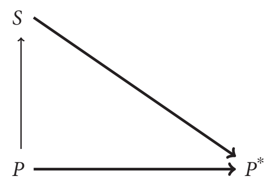

Figure 2.2 Case 2 of the problem of higher-level causation: S causes P^*

So goes Kim's argument that real, distinct, dependent, and efficacious higher-level features induce problematic overdetermination, whether their purported effects are higher-level or lower-level.[8]

Kim sees his argument as motivating rejection of the premise that special-science features are distinct from their base features—that is, he sees it as motivating reductionism (more specifically: reductive physicalism). For present purposes, however (and following Wilson 2011*b* and elsewhere), it is useful to more generally note that rejection of each of the premises of the argument is associated with one or more fairly comprehensive positions in the metaphysics of science. Rejection of one or other of the first four premises gives rise to the following strategies of response, and associated positions:[9]

1. *Substance dualism*. Deny *Dependence*: avoid overdetermination by denying that S and S^* cotemporally materially depend on base features P and P^*, respectively.[10] If higher-level features S and S^* do not so depend on lower-level features, there is no motivation for positing a base feature P as a dependence base for M, hence no motivation for positing a competing causal chain from P to M^* (case 1) or from P to P^* (case 2).
2. *Eliminativism*. Deny *Reality*: avoid overdetermination by denying that S and S^* are real.[11]

[8] Kim also argues that cases of the first type (involving 'same-level' causation) involve cases of the second type (involving 'downward' causation), on grounds that if a given higher-level feature (e.g., S^*) is realized by a lower-level feature (e.g., P^*), any cause of the higher-level feature must also be a cause of the lower-level feature, as per what he calls 'The Causal Realization Principle'; see, e.g., Kim 1993*a*, 439. I will later offer my own reasons for thinking that 'same-level' causation might also involve downward causation, at least in the case of Strong emergence.

[9] In discussing these strategies and associated positions, I flag certain proponents of the positions, for the reader's reference; some such proponents explicitly appeal to the problem of higher-level overdetermination as motivating the position, but not all do so.

[10] See, e.g., Descartes 1641–7/1984 and Chalmers 1996. Pan- or proto-psychists are also appropriately categorized as rejecting *Dependence*, on the assumption (operative here; and reflecting dialectical and historical considerations discussed in Wilson 2006) that (basic) physical goings-on do not individually have or bestow mentality. See, e.g., Seager's (2012) description of a form of panpsychism requiring "that the mental not be ontologically dependent on non-mental features of the world" (19).

[11] See, e.g., Paul Churchland 1981 and 1984, and Patricia Churchland 1986, for eliminativism about certain mental higher-level features, and Merricks 2003 for eliminativism about certain non-mental higher-level features.

3. *Epiphenomenalism*. Deny *Efficacy*: avoid overdetermination by denying that *S* is efficacious.[12]
4. *Reductive physicalism*. Deny *Distinctness*: avoid overdetermination by denying that *S* is distinct from *P*.[13]

Each of these strategies avoids overdetermination, but not in a way that makes sense of higher-level features as metaphysically emergent. In the case at hand, for *S* to be emergent, it must cotemporally materially depend on base feature *P* while being both ontologically and causally autonomous with respect to *P*—that is, while being distinct from and distinctively efficacious with respect to *P*. But the substance dualist strategy denies that *S* cotemporally materially depends on *P*, on grounds that *S* is instantiated in a nonphysical substance.[14] The eliminativist and reductive physicalist strategies each deny that *S* is ontologically autonomous—that is, distinct—from *P*: the eliminativist denies this on grounds that *S* doesn't exist, and the reductive physicalist denies this on grounds that *S* is identical with *P*. Finally, the epiphenomenalist and reductive physicalist strategies each involve denying that *S* is causally autonomous—that is, distinctively efficacious—with respect to *P*: the epiphenomenalist denies this on grounds that *S* isn't efficacious at all, and the reductive physicalist denies this on grounds that *S* and *P* are efficacious in just the same way, since they are identical.

2.1.2 The two 'emergentist' strategies for responding to the problem

The remaining strategies for responding to the problem of higher-level causation, and associated positions, do better by way of accommodating emergence. These are as follows:

5. *Strong emergentism*. Deny *Physical Causal Closure*: avoid overdetermination by denying that every lower-level physical effect has a sufficient purely lower-level physical cause.
6. *Weak emergentism/Nonreductive physicalism*. Deny *Non-overdetermination*: allow that effects caused by *S* are overdetermined by *P*, but maintain that the overdetermination here is of an unproblematic *non*-double-rock-throw variety.

[12] See, e.g., Hodgeson 1962 and Huxley 1874; see Robinson 2012 for contemporary literature.
[13] See, e.g., Smart 1958, Lewis 1966, and Kim 1993*a*.
[14] Similarly, the panpsychist will deny that *S* cotemporally materially depends on physical *P*, on the usual assumption that the (basic) physical goings-on do not themselves have or bestow mentality.

As I argue in the next two sections, these two strategies and associated positions are perspicuously seen as motivated by two conditions on the powers of a given special-science feature, where satisfaction of one or other condition provides a prima facie plausible and principled (i.e., appropriate and illuminating) basis for taking the feature to be emergent, in ways that standard proponents of the strategy/position would endorse. In each of these sections, treating Strong emergence and Weak emergence, respectively, I start by motivating the associated condition on powers by attention to standard versions of the position; I then show how satisfaction of the condition dovetails with the associated strategy for responding to the problem of higher-level causation; I then provide prima facie reasons for thinking that satisfaction of the condition provides an appropriate and illuminating basis for taking special-science features to be both cotemporally materially dependent and ontologically and causally autonomous; finally, I use the condition to formulate the associated schema for metaphysical emergence.

Before getting started, three points of clarification are worth noting. First, as prefigured in Ch. 1 (§1.4.4), talk of 'powers' in what follows is simply shorthand for talk of what causal contributions possession of a given feature makes (or can make, relative to the same laws of nature) to an entity's bringing about an effect, when in certain circumstances (where the circumstances alone are not up to the task of bringing about that effect). Anyone who accepts that the effects an entity causes (or can cause, relative to the same laws of nature) are in part a function of what features the entity has—effectively, all participants to the present debate—is in position to accept 'powers', in the shorthand, metaphysically neutral and nomologically motivated sense at issue here. Besides commitment to the platitude that what entities can do (cause), relative to the same laws of nature, depends on how they are (what features they have), only one metaphysical condition is required in order to make sense of the powers-based conditions to follow; namely, that one's account of (actual or potential) causal contributions (powers) has resources sufficient to ground the identity (or non-identity) of a token causal contribution associated with a token of a higher-level feature, with a token causal contribution associated with a token of a lower-level feature. Here again, effectively all participants to the debate can make sense of such identity (non-identity) claims as applied to token (actual or potential) causal contributions (token powers).[15] Second, the qualifier

[15] For example, suppose a contingentist categoricalist Humean wants to take a physicalist approach to the problem of higher-level causation, and so aims (as I will expand on §2.3) to identify every token power of a token higher-level feature with a token power of its lower-level base feature. As previously discussed, such a Humean understands powers in terms of actual or potential instances of a (contingent) regularity. Where the aim is to avoid overdetermination, the Humean may suppose, to start, that the (relevant instances of the) regularities overlap, both with respect to the (single) effect, and with respect to the (single) circumstances in which the two token features occur. If the Humean aims to be a reductive physicalist, they may suppose that such overlap motivates identifying the token features at issue, and hence the associated powers. If the Humean aims to be a nonreductive physicalist,

'prima facie' in the previous paragraph reflects that a full defense of the claim that the two conditions on powers serve as the basis for two viable conceptions of metaphysical emergence requires detailed treatment of the sort I'll conduct in Chapters 3 and 4. Third, though it is an interesting question how to more specifically understand the forms of dependence at issue in the different schemas for emergence, entering into these details now would take us too far afield. Hence in formulating the schemas, the condition on dependence is expressed simply as requiring cotemporal material dependence, as different forms of emergence agree. Here again, more detailed discussion will be found in later chapters.

2.2 Strong emergentism and the *New Power Condition*

Strong emergentists maintain that some special-science features are real, cotemporally materially dependent on, distinct from, and distinctively efficacious with respect to their base features. So far, Weak emergentists agree—but the conception of higher-level efficacy at issue in Strong emergentism is, as we have seen, one which denies *Physical Causal Closure*, and is correspondingly incompatible with physicalism. And while different accounts of Strong emergentism emphasize different aspects of this distinctive efficacy as located in fundamentally novel features, laws, effects, forces, interactions, and the like, core and common to these accounts is that Strongly emergent features have fundamentally novel powers—powers to produce effects entailing the violation, in particular, of *Physical Causal Closure*.

This is true, to start, on British Emergentism, as endorsed most systematically by Mill (1843/1973), Alexander (1920), Lewes (1875), and Broad (1925).[16] Again, notwithstanding some variations in emphasis, the core claim that Strong emergence involves a fundamentally novel power is a recognizable running theme in these accounts.[17] Hence in his classic survey, McLaughlin (1992) describes British

they can reject this identification of features, on difference-making or other grounds of the sort to be discussed §2.3. Such a Humean will suppose that attention to broader patterns of regularities can provide a basis for identifying token powers of token features, even when the token features are not themselves identical. Whether reductive or nonreductive, the contingentist categoricalist Humean can make sense of the claim that some, all, or none of the token powers of token features are identical. As I observed in my (2015*b*, 35), this case is like the case of New York: if we can make it (out) here, we can make it (out) anywhere.

[16] Evidently Lewes was the first to use the term 'emergent' to characterize certain higher-level entities and features.

[17] In the case of Alexander, there is some controversy about the content (and consistency) of his view, and relatedly, about whether his intended account is supposed to be a version of Strong or of Weak emergence. Here I follow McLaughlin (1992) in grouping Alexander with other Strong emergentists on the basis of certain natural interpretations of some of his statements. Nothing deep hinges on exactly who is categorized as a Strong or Weak emergentist; what is crucial is that accounts of emergence that

Emergentism as "the doctrine that there are fundamental powers to influence motion associated with types of structures of particles that compose certain chemical, biological, and psychological kinds" (52), where the powers at issue are typically taken to be "powers to generate fundamental forces not generated by any pairs of elementary particles" (71).

For example, Mill, the father of British Emergentism, characterizes emergence (in Book III, Ch. 6 of his 1843/1973, 'On the Composition of Causes') in terms of 'heteropathic' effects of joint causes, which effects cannot, unlike 'homopathic' effects, be understood as the mere aggregative sum of the effects of the causes when operating independently—a conception reflecting the supposition, natural at the time, that a violation of broadly additive composition laws of the sort associated, e.g., with the vector addition of forces, sufficed to indicate a fundamentally novel causal influence and associated law. Mill claims that

> To whatever degree we might imagine our knowledge of the properties of the several ingredients of a living body to be extended and perfected, it is certain that no mere summing up of the separate actions of those elements will ever amount to the action of the living body itself. (Book III, Ch. 6, §1)

On Mill's account, then, emergents are distinctively efficacious in being able to produce novel effects—effects going beyond those that base-level goings-on in additive or aggregative combination are able to produce. And though Mill motivates his view by attention to a distinction between diachronically produced effects, it is (as discussed in Ch. 1, note 7) straightforward and desirable to translate Mill's talk of diachronically produced effects into talk of cotemporally had powers: to say that an effect of a feature of a composite entity is heteropathic is just to say that the feature has a power not had by features of its components when in linear combination. Mill himself moves seamlessly from talk of heteropathic effects to talk of new properties of and laws governing entities capable of causing such effects:

> [W]here the principle of Composition of Causes [. . .] fails [. . .] the concurrence of causes is such as to determine a change in the properties of the body generally, and render it subject to new laws, more or less dissimilar to those to which it conformed in its previous state. (1843/1973, 435)

As McLaughlin (1992) notes, "Mill holds that collocations of agents can possess fundamental force-giving properties" (65).

are plausibly seen as compatible with physicalism, or as not so compatible, encode, one way or another, certain identifiable conditions on powers of emergent and base features.

Both Mill's reference to "new laws" and his supposition that such cases contrast with "the extensive and important class of phenomena commonly called mechanical" indicate that Mill takes heteropathic effects to be associated with novel powers, enabling the feature (or its possessing "body") to override the usual composition laws in the production of certain effects. Moreover, as Mill's emphasis on the contrast with effects obeying the 'Composition of Causes' attests, the novel powers at issue in this variety of emergence are *fundamentally* novel, rather than merely reflecting a nonfundamental summation or other aggregation of existing lower-level powers.[18]

For a second example: in his (1920), Alexander (at least on one common reading; see McLaughlin 1992 for discussion) similarly takes fundamentally novel powers to be core to emergence. Consider the following passage:

> Physical and chemical processes of a certain complexity have the quality of life. [. . .] The higher quality emerges from the lower level of existence and has its roots therein, but it emerges therefrom, and it does not belong to that level, but constitutes its possessor a new order of existent with its special laws of behaviour. The existence of emergent qualities thus described is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I should prefer to say in less harsh terms, to be accepted with the "natural piety" of the investigator. It admits no explanation. (46–7)

Here Alexander takes an emergent feature to be indicative of a "new order of existent" with "special laws of behaviour", which novel behaviors are associated with novel powers, as per the usual reciprocal relation between powers and causes. Moreover, Alexander's famous claim that emergent features must be accepted with "natural piety" is strongly suggestive of—or in any case compatible with—a view on which the novel powers at issue are *fundamentally* novel.

Broad (1925) also focuses on laws, in a way that is yet more explicit in the supposition that Strong emergence involves fundamental novelty of powers. In particular, Broad maintains that emergent features arise in accord with funda-

[18] Indeed (and echoing remarks in the Ch. 1, §1.4.2 discussion of the individuation of levels), any powers of a configuration that are novel merely in reflecting a nonfundamental summation or other aggregation of powers of individual components of the configuration are irrelevant to metaphysical emergence of either Strong or Weak varieties, for all parties to the present debate, including reductive physicalists, will happily agree that configurations (pluralities or structural aggregates) can and indeed typically do have such nonfundamentally novel powers. Correspondingly, metaphysical emergence, whatever else it may be, is not a matter of such nonfundamental (i.e., merely aggregative) novelty of powers. At issue in Strong emergence are powers that are fundamentally novel, in being over and above any merely aggregative novelty associated with the coming-to-be of base-level micro-configurations and features, and (as we will shortly see) in Weak emergence no novelty of powers, fundamental or nonfundamental, is at issue.

mental 'configurational' or 'trans-ordinal' laws, connecting lower-level structures with higher-level entities and features, which laws are just as metaphysically and scientifically fundamental as the 'intra-ordinal' laws governing lower-level physical phenomena:

> [T]he law connecting the properties of silver-chloride with those of silver and of chlorine and with the structure of the compound is, so far as we know, an unique and ultimate law. (64–5)

Broad's reference here to emergent laws' being "unique and ultimate" again indicates that the novelty of emergent laws is intended to be of the fundamental rather than nonfundamental variety; and given that in Broad's discussion of candidate cases of emergence the laws at issue are again laws of behavior and associated effects and powers, it follows that core to his account is that Strongly emergent features have fundamentally novel powers.[19]

Contemporary accounts of Strong emergence also typically agree in taking emergent features to have or bestow fundamentally novel powers, not had (or had only in derivative fashion) by base features or associated micro-configurations.[20] For example, Silberstein and McGeever (1999) speak of 'fundamental' irreducibility, and understand emergent features as having causal capacities irreducible to (that is, not identical to) any capacities of features of the parts when in relation:

> Ontologically emergent features are features of systems or wholes that possess causal capacities not reducible to any of the intrinsic causal capacities of the parts nor to any of the (reducible) relations between the parts. (186)

Van Gulick (2001) similarly characterizes what he calls "radical" emergence, highlighting that the novel powers at issue would undercut the physicalist assumption of *Physical Causal Closure*:

[19] Broad put his official account of emergence in terms of 'in-principle' failure of deducibility, but this reflected his assuming that such failures were indicative of fundamental novelty—that is, of Strong emergence. As McLaughlin (1992) notes, "the Emergentists do not maintain that something is an emergent because it is unpredictable. Rather, they maintain that something can be unpredictable because it is an emergent" (73). I'll return to this issue in Ch. 4.

[20] An interesting variation on this theme is found in Paolini Paoletti 2017, Ch. 7, and 2018*a*. Paolini Paoletti suggests that it is in the first instance powers which are Strongly emergent; by way of filling in his schematic (2018*b*) account of emergence as involving partial but not complete dependence, he takes the dependence at issue to reflect that the *having* of the power by a higher-level entity is determined by lower-level powers, whereas the autonomy at issue reflects its being the case that no lower-level goings-on can *manifest* the power. My own view is that there is insufficient daylight between the bestowal and the manifestation of a given power to quite capture the notion of Strong emergence, but in any case Paolini Paoletti's suggestion is a recognizable cousin of the more standard accounts I will discuss here.

> If [. . .] system-level powers were not determined by the laws governing the powers of their parts, then that would seem to imply the existence of powers that could override or violate the laws governing the powers of the parts [. . .] It is in this respect that radically emergent powers would pose such a direct challenge to physicalism, since they would threaten the view of the physical world as a closed causal system. (18–19)

O'Connor and Wong (2005) characterize emergent features as "fundamentally new", not just in being (perhaps epiphenomenally) different, but more specifically in having fundamentally novel causal capacities:

> [A]s a fundamentally new kind of feature, [an emergent feature] will confer causal capacities on the object that go beyond the summation of capacities directly conferred by the object's microstructure. (665)

Hendry (2010) offers an account of Strong emergence continuous with that associated with the British Emergentist tradition, as involving 'configurational Hamiltonians', and similarly maintains that "the emergentist sees [special-science properties] as distinct and nonreducible just because the causal powers they confer are not exhausted by those conferred by their physical bases" (211), contra views committed to "the completeness of physics". More recently, Hendry (2017) maintains that "the conferment of additional causal powers [is] a mark of the distinct reality that is characteristic of something that is strongly emergent" (147).

Finally, in Wilson 1999, I maintain that "the [Strong] emergentist response to the possibility of [mental] causal overdetermination is to deny that the physical is causally closed [. . .] and [assert] that mental properties have causal powers not possessed by any physicalistically acceptable base properties" (41). In Wilson 2002*a*, I go on to argue that, reflecting that scientific practice suggests that powers are plausibly grounded, one way or another, in fundamental forces or interactions (as when the power of a magnet to attract a pin is grounded in the electromagnetic interaction), naturalistic good sense can be made of the Strong emergentist posit of fundamentally novel powers, as reflecting novel fundamental interactions that come into play only at certain levels of compositional complexity, such that Strong emergentism "is committed to there being at least one other fundamental force beyond those fundamental forces currently posited" (74).

2.2.1 The *New Power Condition*

Summing up, Strong emergentists suppose that at least some cotemporally materially dependent features have powers ("causal capacities", etc.) not had by

("not directly conferred by") their (lower-level, physical) base features, as per the following *New Power Condition*:

> *New Power Condition*: Token feature *S* has, on a given occasion, at least one token power not identical with any token power of the token feature *P* upon which *S* cotemporally materially depends, on that occasion.

In having a power that its base feature does not have, *S* has a novel token power. Moreover, and notwithstanding that (as just substantiated), the novel powers at issue in Strong emergence are *fundamentally* novel (as opposed to being non-fundamentally novel, reflecting merely the difference between unaggregated and aggregated lower-level entities and features), there is no need to explicitly register this qualification in the *New Power Condition*. For the condition as stated blocks this sort of nonfundamental novelty, since the condition characterizes the emergence of *S* in terms of a relation between *S*'s powers and the powers of a single lower-level feature *P* (had by a lower-level configuration) upon which *S* cotemporally materially depends, rather than as a relation between *S*'s powers and the powers of multiple lower-level features (had, e.g., by individual components of the entities entering into the configuration having *P*). In other words: if *S* has a power not identical to any power of *P*, where *P* is a feature of a configuration (plurality or structural aggregate) of lower-level (ultimately physical) entities, there is no chance that the novelty of *S*'s power might merely reflect the difference between aggregated and nonaggregated lower-level goings-on; for *P*'s powers will include any powers associated with such aggregative phenomena. This clarification made, here and throughout, the *New Power Condition* should be read as involving a fundamentally rather than a merely nonfundamentally novel power.

Moreover, Strong emergentists typically assume that among the (fundamentally) novel powers of a Strongly emergent feature are powers to influence lower-level physical goings-on, in a way entailing the rejection of *Physical Causal Closure* and ergo the falsity of physicalism. Hence the Strong emergentist position is naturally associated with a strategy of response to the problem of higher-level causation that proceeds by rejecting *Closure*. Moreover, as we'll shortly see, only against the backdrop of the rejection of *Closure* does satisfaction of the *New Power Condition* provide a basis for avoiding overdetermination of higher-level as well as lower-level effects.

The Strong emergentist response to the problem of higher-level causation

Let's now return to the problem of higher-level causation, starting with the second, simpler case, to see how satisfaction of the *New Power Condition* enters into implementation of the Strong emergentist strategy for responding to this problem.

In the case where special-science feature S causes a base feature P^* (case 2), the Strong emergentist strategy involves, to start, the supposition that S satisfies the *New Power Condition* specifically in having a fundamentally novel power to bring about P^*. For example, S might be a Strongly emergent state of being thirsty, which depends on base feature P, and which in the circumstances causes a physical reaching for a nearby glass of water P^*. On this assumption, P^* does not, contrary to the assumption of *Physical Causal Closure*, have a sufficient purely lower-level physical cause: as per the *New Power Condition*, P has no token power identical with S's token power to cause P^*; hence either P is not at all a cause of P^* (does not have any power to cause P^*), or else, if P can be understood to cause P^* (that is, to have a power to cause P^*), P has this power only in a derivative way, in virtue of P's being a dependence base for S, which non-derivatively has the power at issue.[21] Either way, P fails to be a sufficient purely lower-level physical cause of P^*; and without loss of generality, it moreover follows that P^* has no sufficient purely lower-level physical cause, contra *Physical Causal Closure*, and overdetermination is avoided, as in Figure 2.3:

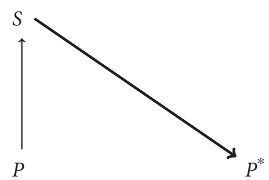

Figure 2.3 The Strong emergentist's response to case 2

Next, suppose (as per case 1), that S causes another special-science feature S^*—say, a desire to drink some water. Here the Strong emergentist supposition is that S satisfies the *New Power Condition* specifically in having a fundamentally novel power to bring about S^*—that is, a power that P doesn't have (either at all, or non-derivatively). Interestingly, even though the novel power at issue here is not directed at the production of a lower-level physical effect, it remains that satisfaction of the *New Power Condition* in this case requires the falsity of *Physical Causal Closure*. Why so? Because, if *Closure* held in this case, P *would* have a non-derivative power to cause S^*—by being a sufficient purely lower-level physical cause of P^*, which in turn nomologically necessitates S^*. But in that case overdetermination would not be avoided, and moreover the claim that S has a fundamentally novel power to cause S^* would be undermined. Hence the Strong emergentist must deny *Physical Causal Closure*, even when the novel power had by

[21] S's causing of P^* might be entirely independent of P, or it might be that S and P jointly cause P^*; either route to the production of P^* is compatible with the denial of *Physical Causal Closure*. I'll revisit these options down the line.

Strongly emergent *S* is for the causing of a special-science feature *S**. In this case, the strategy is as in Figure 2.4:

$$\begin{tikzcd} S \arrow[r] & S^* \\ P \arrow[u] & P^* \arrow[u] \end{tikzcd}$$

Figure 2.4 The Strong emergentist's response to case 1

A remaining question about the Strong emergentist treatment of this case concerns what is responsible for *S**'s having the base feature *P** that it does. Given that *P* is not (on this view) itself up to the task of causing *P**, there are two possibilities here: first is that *S** carries with it its own dependence base *P**, such that *S*, in causing *S**, also causes *P** (such that cases of type 1, in which *S* causes *S**, turn out also to be cases of type 2, in which *S* causes *P**); another is that *S* and *P* jointly cause *P** (with *S* either independently causing *S** or else causing *S** jointly with *P*). Either way, the Strong emergentist response to case 1, like the response to case 2, involves so-called 'downward causation'. I will explore these options in more detail in Ch. 4.

2.2.2 The schema for Strong emergence

Prima facie, satisfaction of the *New Power Condition* by a special-science feature *S* which cotemporally materially depends on a base feature *P* provides an appropriate and illuminating basis for avoiding overdetermination while guaranteeing that *S* is both ontologically and causally autonomous with respect to *P*. First, since *S* has a token power (at a time or over a temporal interval) that *P* doesn't have (at that time or over that interval), *S* is distinct from *P* (by Leibniz's law); hence *S* is ontologically autonomous with respect to *P*. Second, in having a novel token power, *S* can cause an effect that *P* can't cause, or that *P* can't cause in the same (non-derivative) way as *S*; hence *S* is causally autonomous—that is, distinctively efficacious—with respect to *P*. The *New Power Condition* at the heart of the Strong emergentist's strategy for resolving the problem of higher-level causation thus provides the basis for our first schema for metaphysical emergence:

> *Strong Emergence*: What it is for token feature *S* to be Strongly metaphysically emergent from token feature *P* on a given occasion is for it to be the case, on that occasion, (i) that *S* cotemporally materially depends on *P*, and (ii) that *S* has at least one token power not identical with any token power of *P*.

Here the locution 'what it is for' is intended to flag that *Strong Emergence* provides a schematic metaphysical basis for a given case of such emergence, encoding what is core and crucial to that notion. Of course, any schematic account needs to be sensibly filled in. Modulo this caveat, the conditions in the schema are, I claim, necessary and sufficient for (appropriate and illuminating accommodation of) metaphysical emergence of the physically unacceptable variety.[22] Some clarifications:

- The schema is formulated so as to apply to features of artifacts and the like as well as to features of special-science entities.
- The first condition minimally specifies cotemporal material dependence, typically and comparatively neutrally understood as involving both (physical) substance monism and the minimal nomological supervenience of emergent feature types on base feature types.
- The second condition (effectively, the *New Power Condition*) captures the comparatively strong sense in which an emergent feature may be causally, hence ontologically, autonomous with respect to the base feature upon which it cotemporally materially depends, in virtue of having a fundamentally novel power relative to the powers of its base feature.

  Note that while the initial presentation of the *New Power Condition* built in the cotemporal material dependence of the feature having the new power on its base feature, the schema for *Strong Emergence* treats the new power and the dependence conditions separately. Henceforth, references to the *New Power Condition* should be understood as references just to condition (ii) in that schema.
- As I have been emphasizing throughout, the base feature at issue is a feature of a given (ultimately physical micro-) configuration (plurality or structural aggregate), and the novelty of the power at issue in condition (ii) should be understood accordingly. In particular, the novelty of this power is not to be understood as relative to the powers of features of individual entities entering into the dependence base configuration.
- Again, *S*'s novel power is not just nonfundamentally novel (as reflecting a merely aggregative difference between powers of a configuration and powers of components of the configuration) but is moreover fundamentally novel. I do not include the qualifier 'fundamentally' in the schema because including it would misleadingly suggest that the nonfundamentality of *S*'s novel power was among the available options, which (as previously discussed) it isn't, under the operative understanding of *P* as having any nonfundamental powers attaching to mere aggregation of lower-level components.

[22] This is not to presuppose, of course, that all implementations of *Strong Emergence* are on a par or equally well suited to accommodate a given macro-phenomenon. As we'll see in Ch. 4, my preferred implementation has certain advantages over others.

- The schema is relativized to occasions (times or temporal intervals), but it is worth noting that it would be reasonable to suppose that it suffices for the Strong emergence of *S*, *simpliciter*, that the condition is ever satisfied, and to suppose that it suffices for the Strong emergence of the feature type (of which *S* is a token), *simpliciter*, that any token feature *S* on any occasion satisfies (or would satisfy) the condition. These complications won't play a role in what follows.

2.3 Nonreductive physicalism and the *Proper Subset of Powers Condition*

Nonreductive physicalists, like Strong emergentists, maintain that (some) special-science features are real, cotemporally materially dependent, distinct, and distinctively efficacious with respect to their base features. They too are non-reductionists, maintaining that certain configurations (pluralities or structural aggregates) of entities whose constituents are ultimately physical have features from which other features emerge, constituting emergent entities and a novel level of natural reality. But as physicalists, their response to the problem of higher-level causation cannot entail the rejection of *Physical Causal Closure*, which thesis is core to the physicalist view that the physical goings-on are an existential and causal basis for all other broadly scientific phenomena. Rather, nonreductive physicalists reject *Non-overdetermination*, maintaining that distinct special-science and base features can each be a sufficient cause of a single effect, in virtue of standing in a relation that, while not identity, is intimate enough both to avoid overdetermination of the problematic (since implausible, for the cases at issue) double-rock-throw variety and to retain compatibility with *Physical Causal Closure*, hence with physicalism.[23]

In presenting their view, nonreductive physicalists (or those aiming to characterize this position) typically endorse some or other 'realization' relation as holding between tokens or types of the features at issue, which relation, either in itself or in combination with some consonant condition, is supposed to illustrate or illuminate how problematic overdetermination is avoided, compatible with both physicalism and nonreduction, as in these representative cases:

- Functional realization (Putnam 1967, Fodor 1974, Papineau 1993, Antony and Levine 1997, Melnyk 2003, Witmer 2003, Polger 2007, Yates 2012): realized features are functionally implemented by lower-level physically acceptable features. As applied to mental properties:

[23] How nonreductive physicalists aim to establish that higher-level features are not just efficacious, but distinctively so, will be addressed shortly.

> Let "functionalism" be the view that, first, each mental property *M* is identical with the second-order property of having some property or other which plays the causal role associated with *M*, and second, each actual instance of a mental property is realized by some physical property. If anything counts as a received theory of the mental in contemporary philosophy of mind, this is it. (Witmer 2003, 198)

> [T]he fact that psychological properties are realized in some physical property or other seemed to ensure that there was nothing occult or dualistic about psychological causation: the causal interactions among the inputs, outputs, and internal states of a device that realized a higher-order functional system would be effected by the causal mechanisms characteristic of the first-order realizers.
> (Antony and Levine 1997, 86)

- Constitutive mechanism (Cummins 1975, Craver 2001, Haug 2010): realized features of a system are implemented by lower-level micro-structural features specifying causal mechanisms involving parts of the system:

> The components of mechanisms, in contrast to those of mere aggregates, have an active organization; they act and interact with one another in such a way that the ψ-ing of [realized feature] *S* is more than just the sum of [individual component] φ properties. (Craver 2001, 59)

> [T]he realizer *P* must itself provide a causal mechanism for the instantiation of *M*—a constitutive mechanism, in which the instantiation of *M* in *s* is explained by the components of *s* being propertied and related in certain ways. (Haug 2010, 321)

- Mereological realization (Shoemaker 2000/2001, Clapp 2001, Rueger and McGivern 2010): realized features are proper parts of lower-level physical features:

> [T]he instantiation of a realizer property entails, and might be naturally said to include as a part, the instantiation of the [. . .] property realized.
> (Shoemaker 2000/2001, 28)

> [M]ultiply realized mental properties, though real and causally efficacious, are better thought of as *parts* of their physical realizers. [. . .] Just as there is no causal and/or explanatory competition between a whole and its parts, so there is no causal and/or explanatory competition between instances of mental properties and instances of their physical realizers. (Clapp 2001, 133; italics in text)

- The determinable/determinate relation (Macdonald and Macdonald 1986, Yablo 1992, Wilson 1999 and 2009): realized features are determinables of lower-level physical features. Again as applied to mental phenomena:

  > Traditionally, the paradigm of one-way necessitation was the relation of determinate to determinable [. . .]. What if mental phenomena are determinables of physical phenomena in something like the traditional sense [. . .]? (Yablo 1992, 250)

  > [W]e know that [determinables and determinates] are not causal rivals. This kind of position is of course familiar from other contexts. Take for example the claim that a space completely filled by one object can contain no other. Then are even the object's parts crowded out? No. In this competition wholes and parts are not on opposing teams [. . .].
  > (Yablo 1992, 183)

- 'Superdupervenience' or explanatory necessary connection (Lepore and Loewer 1987 and 1989, Horgan 1989 and 1993). As Horgan (1993) describes the view:

  > [T]he sort of inter-level relation needed by the materialist who is also a realist about a given mode of discourse (e.g., mental discourse) is not bare supervenience, but rather what I hereby dub superdupervenience: viz., ontological supervenience that is robustly explainable in a materialistically explainable way. (566)

  Here the unproblematic nature of the overdetermination at issue is reflected in the association of the schematically characterized realization relation with an additional, broadly counterfactual condition encoding the causal relevance—or as Horgan evocatively puts it, 'quausal' relevance, whereby a feature is efficacious '*qua*' the type of feature it is—of the realized feature:

  > What seems required for quausal relevance [of a property F vis-à-vis an effect having property G], then, is a wider pattern of counterfactual dependence of the occurrence and non-occurrence of events with property G upon the occurrence or non-occurrence of events with property F. (Horgan 1989, 59)

  > [S]ubjunctive conditionals [. . .] lie at the heart of the view that [mental] content properties are causally potent.
  > (Lepore and Loewer 1989, 188)

Now, underlying the seeming diversity in these and many other accounts of nonreductive physicalism hides a deeper unity of strategy, which again can be put in terms of a certain condition on powers (see Wilson 1999 and 2011*b*). Motivating this condition—the *Proper Subset of Powers Condition*—and seeing how it provides

a unifying basis for a wide range of nonreductive physicalist responses to the problem of higher-level causation takes a bit more doing, however, than in the corresponding case of the *New Power Condition* as operative in accounts of Strong emergence. I'll start by putting the *Proper Subset of Powers Condition* on the table, and arguing that each of the aforementioned accounts of realization aims to ensure its satisfaction; I'll then argue that satisfaction of this condition provides a basis not just for a cotemporally materially dependent feature *S* to be distinct, but also for *S* to be distinctively efficacious; I'll then show how this condition provides an illuminating basis for understanding the nonreductive physicalist's response to the problem of higher-level causation; finally, I'll use the condition as the basis for the schema for Weak emergence.

2.3.1 The *Proper Subset of Powers Condition*

Again, physicalists, unlike Strong emergentists, cannot allow that special-scientific features have powers different from the physical base features upon which they cotemporally materially depend, on a given occasion; for if some special-scientific feature *S* were to have a power its physical base feature didn't have, then this would violate the physicalist supposition that the physical goings-on provide a complete ontological and causal basis for all other broadly scientific goings-on. Correspondingly, physicalists are committed to every special-science feature *S*'s satisfying the following condition:

> *Token Identity of Powers Condition*: Every token power of token feature *S*, on a given occasion, is identical with a token power of the token feature *P* on which *S* cotemporally materially depends, on that occasion.

Note that *reductive* physicalists can and do accept the *Token Identity of Powers Condition*, so long as it is allowed that a feature can depend on itself in the relevant sense. For example, if a given mental type is (as the reductive physicalist believes) identical to a given physical type, then tokens of these types, along with their associated token powers, will also be identical, and the *Token Identity of Powers Condition* will be satisfied.

As I'll now argue by attention to the aforementioned accounts of realization, nonreductive physicalists moreover typically accept another, stronger condition, according to which at least some higher-level features *S* stand in a relation to their lower-level base features *P* satisfying the following condition:

> *Proper Subset of Powers Condition*: Token feature *S* has, on a given occasion, a non-empty proper subset of the token powers of the token feature *P* on which *S* cotemporally materially depends, on that occasion.[24]

Functional realization

First, consider functionalist accounts of realization, on which realized types are higher-order types associated with causal roles that, on a given occasion, are played by tokens of realizer types. A causal role is just a collection of powers. Hence if *S* is a token feature of functional type, then, on any given occasion, every token power of *S* will be numerically identical with a power of the base feature *P* that plays *S*'s causal role on that occasion. Witmer (2003) concurs, saying that on a functionalist approach, "any effect—not just physical ones—to which a mental property may seem causally related is one which can be explained by reference to the causal powers of the realizer property" (200).[25] Functionalist accounts of realization thus satisfy the *Token Identity of Powers Condition*. Do such accounts also satisfy the *Proper Subset of Powers Condition*? It is sometimes suggested that instances of functionally realized features inherit *all* of the token powers of their realizing feature instances:

> A functional reduction of pain has the following causal and ontological implications: Each occurrence of pain has the causal powers of its neural realizer [. . .] In general, if *M* occurs by being realized by *N* on a given occasion, the *M*-instance has the causal powers of the *N*-instance.
>
> (Kim 2006, 554)

In cases of multiple realizability, however, a functionally realized feature arguably has only a proper subset of the powers of its realizing feature(s), at both the type and token levels.

To see this, first recall the hardware/software analogy motivating functionalism, initially highlighted by Putnam (1967): the realizing systems are similar with respect to powers needed to implement a given software program, but different with respect to powers associated with their distinctive varieties of hardware. More generally, when a type of functionally characterized feature is multiply realizable, its realizing types will each have all the powers associated with the functional role, and more besides (where the further powers reflect differences between the

[24] The requirement that the proper subset of powers be non-empty reflects the rejection of epiphenomenal features as metaphysically emergent, in the relevant sense.

[25] See also Yates 2016: "According to functionalism, mental properties are defined by roles that are filled by physical properties, so a novel causal role seems out of the question" (4); "Functional properties are physically realized, and their realizers do all the causal work associated with their defining roles" (11).

multiple realizers). Hence the powers of the realized type will be a proper subset of those of each of its realizing types.

Moreover, this proper subset relation between powers of the types will arguably hold between token powers of instantiations of the types, as per the *Proper Subset of Powers Condition*. It may make sense for a token feature to have fewer powers than its feature type, reflecting restrictions associated with circumstances in which the feature occurs or is instantiated.[26] But as I've previously discussed (see Wilson 2011*b* and 2015*b*), it makes no sense for a token feature, whether functionally realized or not, to have more powers than its type; for if a token feature purportedly of a certain type had token powers not associated with the purported type, that would be reason for taking the token not to be of the type. So functionally realized features arguably satisfy the *Proper Subset of Powers Condition*.[27]

Constitutive mechanism realization

Second, consider constitutive mechanism accounts of realization. Such accounts trace back at least to the discussions in Cummins 1975 and 1983, which in turn draw on explanatory strategies in the sciences aimed at "understanding the capacities of systems by analyzing them into the capacities of their components" (Craver 2001, 54), as in the following classic case of a mechanistic explanation:

> [T]he circulatory system's (S's) capacity to deliver goods to body tissues (W) is explained by decomposing it into its parts (e.g., hearts (X_1), arteries (X_2), kidneys (X_3), and valves (X_4)) and capacities (e.g., to pump (φ_1), to convey (φ_2), to filter (φ_3), and to regulate the direction of blood flow (φ_4)) and linking those parts together in the programmed φ-ing of the circulatory system.
> (Craver 2001, 55–6)

[26] Clarke (1999, 307) suggests something in this ballpark, albeit as applied to what he calls 'exemplifications' of properties in events as opposed to token instantiations of properties.

[27] The preceding point serves as my primary response to Morris's (2013) claim that someone maintaining that instances of higher-level features inherit *all* of the token powers of their realizers, as per what he calls 'Full Inheritance', can accommodate the presumption that tokens of a multiply realized property have the same powers regardless of how they are realized. As Morris puts it:

> [T]here is a sense in which the defender of Full Inheritance can accept the causal unity of multiply realized properties—namely, in the sense that all instances of a property *M* have a certain set of causal powers. A functionalist about pain who endorses Full Inheritance may, for example, allow that all instances of pain have the power to bring about avoidance behavior. It is just that the defender of Full Inheritance will then contend that insofar as pain is multiply realized, there may be many other powers that are possessed by some instances of pain but not others. (207)

But as noted in the text, that some token feature has powers not associated with a given type is good reason to think that the feature is not of that type. Coupled with the usual supposition that functional types are associated with comparatively abstract roles, this observation provides independent good reason to endorse Subset rather than Full inheritance as characteristic of functionally realized features.

Such explanatory strategies have inspired an approach to realization of higher-level features of a system explicitly encoding reference to the parts of the system and the specific ways (associated, e.g., with individual 'role functions') in which the parts and their features enter into implementing the higher-level capacity or function.[28] On this approach, a mechanistically realized feature is realized by a single lower-level feature—sometimes called a 'micro-structural' feature—schematically along lines of *the feature of having such-and-such spatiotemporally related parts engaging in such-and-such causal interactions* (exercising such-and-such capacities, engaging in such-and-such activities). Such an approach is suggested by Craver's (2001) observation that precisely what is distinctive about mechanistic approaches is that the parts are engaging in organized as opposed to merely aggregative behavior: "The components of mechanisms, in contrast to those of mere aggregates, have an active organization; they act and interact with one another in such a way that the ψ-ing of [realized feature] S is more than just the sum of [individual component] φ properties" (59). Kim describes such a mechanistic account of realization in such terms, as follows:

> When P is said to 'realize' M in system s, P must specify a microstructural property of s that provides a causal mechanism for the implementation of M in s [. . .]. (Kim 1993, 197)

As Haug (2010) more recently puts it (expanding on the earlier quotation):

> [T]he realizer P must itself provide a causal mechanism for the instantiation of M—a constitutive mechanism, in which the instantiation of M in s is explained by the components of s being propertied and related in certain ways [. . .] in a constitutive mechanism, the properties [. . .] involved in the mechanism belong to proper parts or sub-systems of s. [. . .] a constitutive mechanism explains how it is that a given special science property is instantiated [. . .] For example, the microstructure of acids (including the crucial activity of proton donation) provides a constitutive mechanism for being an acid and is part of an integrative [i.e., nomological] mechanism that can be used to explain why acids behave the way they do (e.g., dissolve metal, form a salt when combined with a base, turn litmus paper red). (321)

[28] In fact, mechanistic explanations have inspired two different accounts of realization: a one-one approach, to be next treated, and a many-one approach, of the sort developed by Gillett in his 2002*b*, 2002*a*, 2016, and elsewhere. I will treat Gillett's approach in Ch. 3 (§3.4.2).

A constitutive mechanism account of realization along these lines arguably ensures satisfaction of the *Proper Subset Condition on Powers*. One way to see this is to note that such an approach is a clear variation on the theme of functional realization, with the main difference reflecting that mechanistic approaches, in being concerned to identify just *how* a given higher-level functional role is implemented by lower-level goings-on, explicitly require that the realizing feature encode the micro-structural and causal mechanistic means by which the role associated with the higher-level feature is implemented. Notwithstanding this emphasis, it remains that on this view, realized features are associated with causal roles that, on a given occasion, are played by tokens of realizing features, and that a causal role is just a collection of powers. Hence if *S* is a token feature of a mechanistically realized type, then, on any given occasion, every token power of *S* will be numerically identical with a token power of the micro-structural base feature *P* that mechanistically implements *S*'s causal role on that occasion, as per the *Proper Subset Condition on Powers*. And indeed, Haug includes satisfaction of a subset-of-powers condition in his considered mechanistic account:

> I propose the following positive account of the way in which physical properties realize mental (and other high-level) properties. A physical property *X* realizes property *Y* if and only if (a) the set of causal powers contributed by *Y* is a subset of the causal powers contributed by *X* [. . .] and (b) the components of *X* provide a constitutive mechanism for the instantiation of *Y*, which (c) fits into a family of realizers that, in turn, provides an integrative mechanism for *Y*'s causal or nomological relations to other (usually high-level) properties. (323)

As I read Haug, he does not mean here to suggest that satisfaction of conditions (b) or (c) does not entail or ensure satisfaction of the *Proper Subset of Powers Condition* at issue in (a); rather (as his discussion of alternative accounts of realization satisfying the latter condition indicates), he simply aims to make explicit the mechanistic means by which the *Proper Subset of Powers Condition* is met.

Mereological realization

Third, consider mereological (parthood-based) accounts of realization, according to which realized features are proper parts of their realizing features at the level of both types and tokens (Shoemaker 2000/2001 and 2007, and Clapp 2001). Proper parthood appears to satisfy the nonreductive physicalist's desiderata: proper parts are distinct from and yet in a sense completely metaphysically dependent on (grounded in, constituted by, etc.) a whole that is antecedently given, and the parts may be efficacious without inducing overdetermination, as when both I and my eye cause a wink, or both a plane and its wheels are causes of a runway's being

touched (as in Paul's 2002 example). In the present case, the suggestion is that a lower-level physical feature is something like a more fundamental whole of which a higher-level feature is a less fundamental proper part.[29]

Both Shoemaker and Clapp suppose that mereological accounts satisfy the *Proper Subset of Powers Condition*; indeed, both see satisfaction of the condition as core to their accounts of realization, with the appeal to mereology serving to illustrate their preferred strategy for satisfying the condition.

To start, Shoemaker (2000/2001) presents an account of realization based on a type-level version of the *Proper Subset of Powers Condition*:

> Property *X* realizes property *Y* just in case the conditional powers bestowed by *Y* are a subset of the conditional powers bestowed by *X* (and *X* is not a conjunctive property having *Y* as a conjunct). (78)

He then claims that multiply realized feature types satisfy this condition:

> Where the realized property is multiply realizable, the conditional powers bestowed by it will be a proper subset of the sets bestowed by each of the realizer properties. (78–9)

Shoemaker supports this claim by appeal to considerations similar to those canvassed for functional realization, with the main difference being that (following the causal account of properties advanced in his 1980) he takes all broadly scientific properties to be essentially characterized by distinctive sets of powers. When such a feature is multiply realized, he plausibly maintains, its realizing types will share all the powers of the realized type, but will differ from each other in respect of further powers. Shoemaker goes on to argue that this proper subset relation between powers of the types will plausibly hold between token powers of the instantiated types. In general, Shoemaker (2000/2001) notes, it is reasonable to suppose that if realized and realizer types are not identical, in virtue (at least in part) of bestowing different sets of conditional powers, then neither will be their instantiations:

> [I]t seems doubtful that we should identify the mental property instance with the instance of the physical property that realizes it—or that we should identify the instance of red and the instance of scarlet. If we think of the instantiation of a property as the conferring on something of the conditional

[29] It is important to this strategy that the whole be antecedently given as (more) fundamental. Were the parts to be antecedently given as (more) fundamental, it might rather be the whole that was appropriately seen as completely metaphysically dependent on the parts. See Wilson 2014*a* for further discussion.

> powers associated with that property, then when properties confer different sets of conditional powers, the instantiation of one of them is not identical with the instantiation of the other. (28)

These remarks suggest that in cases of multiple mereological realization, and given that one endorses a causal account of properties, the *Proper Subset of Powers Condition* is satisfied.[30]

Alternatively, one can backwards-engineer the need to satisfy the *Proper Subset of Powers Condition* from a mereological approach, independently of whether one endorses a causal account of properties. In particular: if features have non-causal aspects, then the supposition that a realized state is a proper part of a realizing state need not indicate that the realized state has any powers at all (if the overlap concerns only the noncausal aspect), much less that it is distinctively efficacious. Hence if proper parthood is to provide a basis for higher-level efficacy, the overlap must be specifically in respect of powers, as per the *Proper Subset of Powers Condition*.[31]

Determinable-based realization

Fourth, consider accounts of realization in terms of the determinable/determinate relation, the relation of increased specificity paradigmatically holding between (determinable) colors and their (more determinate) shades (see Macdonald and Macdonald 1986 and 1995, Yablo 1992, Ehring 1996, Wilson 2009). Yablo (1992) expected the suggestion that, e.g., mental features stand to their physical realizations in the relation that colors bear to their shades to be met with some incredulity. One way to make his conjecture more plausible is to put the point in terms of the causal powers of the properties involved (see Wilson 1999 and 2009). Consider a patch that is red, and more specifically scarlet. Sophie the pigeon, trained to peck at any red patch, is presented with the patch, and she pecks. The patch's being red caused Sophie to peck—after all, she was trained to peck at red patches. But the patch's being scarlet also caused Sophie to peck—after all, to be scarlet just is to

[30] Similarly for the mereological approach presented in Rueger and McGivern 2010, on which higher-level behaviors are proper parts of lower-level behaviors: "[W]e can understand the claim about parthood between behaviours in terms of the subset relation between their associated causal powers—so that the causal powers of the macro behaviour form a subset of those of the micro behaviour [...]" (395).

[31] Importantly, and notwithstanding that Shoemaker and Clapp each take a mereological account of realization to naturally flow from a causal theory of properties, such an understanding of properties is not required in order for an account of realization to be seen as ensuring satisfaction of the *Proper Subset of Powers Condition*. I'll return to this issue in Ch. 3, when considering the objection (pressed by Melnyk 2006, among others) that satisfaction of the conditions in the schema for Weak emergence will guarantee the physical acceptability of the associated emergent feature only if a causal theory of properties is assumed.

be red, in a specific way. Nonetheless, Sophie's pecking was not problematically overdetermined. Plausibly, this is because each token power of the determinable red instance is numerically identical to a token power of its determining scarlet instance. Similarly, the proponent of a determinable-based account of realization maintains, for the case of *S* and *P*, in which case such an account satisfies the *Token Identity of Powers Condition*.

Here too it has been suggested that instances of determinable features inherit *all* of the token powers of their realizing feature instances, on grounds that determinable and determinate instances are identical (Macdonald and Macdonald 1995, Ehring 1996). But again, there is a case to be made that the powers of a determinable feature are a proper subset of those of its determinate features, at both the type and the token levels.

To start, note that given Sophie's training, she would have pecked even had the patch been a different shade of red (burgundy, say); but not so for Sophie's cousin Alice, trained to peck only at scarlet patches. This suggests that the determinable feature type *red* has fewer powers than its determinate types (*scarlet*, *crimson*, etc.). More generally, since broadly scientific determinables are associated (as per usual: at least in worlds with the same laws of nature) with distinctive sets of powers, and may typically be determined by multiple more specific determinates, the powers of determinable feature types will typically be a proper subset of those of their determinate feature types. Moreover, this relation will plausibly hold between token powers of determinable and determinate instances. Again, while it might make sense for an instance of a given feature type to have fewer powers than are associated with its type (reflecting restrictions on available circumstances of instantiation, or the like), it does not make sense for an instance of a given feature type to have more powers than are associated with the type. In particular, were a token feature purportedly of a determinable type to have more token powers than are associated with the type, that would be reason to think that the token was not, after all, of that type. Hence a determinable/determinate account of realization arguably satisfies the *Proper Subset of Powers Condition*.

Ontologically explanatory realization

Finally, consider the aforementioned schematic accounts of realization as involving an ontologically explanatory necessary connection; here I focus on that offered in Lepore and Loewer 1989, which draws on Lepore and Loewer 1987 and Horgan 1989. Lepore and Loewer sketch a strategy whereby a feature that is so realized might be distinctively efficacious with respect to its base feature in virtue of certain subjunctive conditionals being true:[32]

[32] In the quotations to follow, I've corrected Lepore and Loewer's spelling of 'quausation' for consistency with Horgan's discussion.

> When explaining why one thinks that possession of a [higher-level] property makes a causal difference one is likely to make reference to certain subjunctive conditionals. [. . .] We suggest that [such] conditionals lie at the heart of the view that [higher-level] properties are causally potent. To apply this idea to characterizing causal potency we will define a relation between event property pairs which, after Terry Horgan (1989), we call "quausation" (as in *c qua F* causes *e qua G*). As a tentative account of the quausation relation we propose the following:
>
> > (*c*, *F*) is quausally related to (*e*, *G*) iff *c* and *e* occur and are respectively *F* and *G* and there is some time before the occurrence of *c* at which these two conditionals obtain: (1) if *c* were to occur and be *F* then that would cause an event *e* to be *G*; (2) if *c* were to occur but not be an *F* then it would not cause an event which is *G*. (188–9)

Crucially, the proposal eschews the suggestion that distinctive efficacy requires new powers, on grounds that this route to such efficacy would be at odds with physicalism. Rather, the distinctive efficacy at issue reflects the sense in which higher-level features make a difference to whether the effect occurs:

> As far as we can see an event's causal powers are completely determined by its basic causal properties. [Higher-level] properties are not needed for that. However, if (*c*, *F*) is quausally related to (*e*, *G*) then there is a perfectly good sense in which *c*'s having *F* makes a difference to *c*'s causal powers. This is the sense captured by the counterfactuals 1 and 2. To say that instantiating *F* makes a difference to what *c* causes is to say, in part, that had *c* not had *F* then it would not have caused a *G*. Of course, this counterfactual is true in virtue of certain physical facts obtaining. But that doesn't make these conditionals any less true or any less explanatory. (190)

For reasons similar to those associated with a determinable-based approach to realization, a counterfactual or difference-making approach to distinctive higher-level efficacy is arguably one entailing satisfaction of the *Proper Subset of Powers Condition*.

Summing up: a representative range of seemingly diverse accounts of realization are unified in each arguably ensuring satisfaction of the *Proper Subset of Powers Condition* as a means of avoiding problematic overdetermination. Other accounts of realization arguably ensure satisfaction of this condition, as well—for example (as I argue in Wilson 1999), Pettit's (1995) account as appealing to a 'dot-shape' analogy, according to which the shape is distinct from, but nothing over and above, a collection of dots, and (as I argue in Wilson 2010*b* and will

discuss in detail in Ch. 5) my account in terms of an elimination in degrees of freedom.

Distinctive power profiles as a basis for ontological and causal autonomy

In having only a proper subset of the token powers of the token physical feature *P* upon which it depends, on a given occasion, a token feature *S* satisfying the *Proper Subset of Powers Condition* will clearly be distinct from *P*, by Leibniz's Law. Might *S* also be causally autonomous—distinctively efficacious—with respect to *P*, as required if *S* is to be genuinely metaphysically emergent?

Yes, supposing that a feature may be distinctively efficacious in virtue of having a distinctive *set or collection* of causal powers—i.e., a distinctive power profile. The underlying promise of nonreductive physicalism as making sense of the distinctive efficacy of (at least some) special-science (more generally, higher-level) entities and features lies in the claim that *S*'s causal autonomy does not require that *S* have a distinctive power: it is enough that *S* have a distinctive set (collection, plurality) of powers.

One case for taking the having of a distinctive power profile to be sufficient for causal autonomy adverts to the sort of difference-making or other 'proportionality' considerations operative in cases where *S* (or *S*'s type) is multiply realizable (following discussion in, e.g., Lepore and Loewer 1987 and 1989, Horgan 1989, and Yablo 1992).[33] Again, suppose that *S* is a state of feeling thirsty, which causes a physical reaching for a glass of water (effect *E*). Now suppose that *S* (or another token of *S*'s type, etc.) were realized by *P′* rather than *P*, in circumstances relevantly similar to those in which *S* caused *E*. Would *E* (or an event of *E*'s type) have still occurred? Intuitively, yes, since the only powers that matter for the production of *E* are the powers associated with *S*: powers differing between *P* and *P′* (e.g., to produce different readings on a neuron detector) don't make a difference to, and hence are in this sense irrelevant for, *E*'s production. That *S*'s distinctive power profile contains just those powers relevant or 'proportional' to *E*'s production provides a principled reason for taking *S*'s efficacy vis-à-vis *E* to be distinctively different from *P*'s, notwithstanding that (as per *S*'s satisfaction of the *Token Identity*

[33] Yablo suggests that a candidate determinable cause (e.g., the patch's being red) might be more proportional to a given effect (e.g., Sophie's pecking), on a given occasion, than the associated candidate determinate cause (e.g., the patch's being scarlet), in that the determinable has an 'essence' tracking both sufficiency and difference-making considerations (e.g., if the patch had been crimson rather than scarlet, Sophie would still have pecked). Here I focus on difference-making considerations, since (against the background assumption of *Physical Causal Closure* and associated satisfaction of the *Token Identity of Powers* condition, in particular) it is difference-making rather than sufficiency which is distinctive of higher-level efficacy, on a 'proper subset of powers' approach to realization.

and *Proper Subset* conditions on token powers), *S* doesn't cause anything that *P* (or other realizers of *S*'s type, on other occasions) doesn't (or don't) also cause.[34]

Another case for causal autonomy reflects that distinctive power profiles are typically associated with distinctive systems of laws—for example, the special-science laws treating entities of *S*'s type. Plausibly, systems of laws track causal joints in nature; hence when *S* is of a special-science type, its distinctive power profile is similarly plausibly understood as tracking such a distinctive causal joint. Moreover, given the holding of the *Proper Subset of Powers Condition*, and consonant with a general understanding of special-science phenomena as abstracting away from various lower-level details, the causal joints at issue here concern comparatively abstract goings-on. This is, I think, what Antony and Levine (1997) have in mind when they say that for the causal autonomy of functional properties, "What we really need is a 'realization-indifferent' regularity: a contingent regularity that essentially involves the [realized] property, and that applies to any instance of the property, no matter the form of realization" (92). Moreover, as Antony (2003) notes, even in the absence of multiple realizers, one can still make sense of the presence of properties and laws that are, as she puts it "*at a higher level of abstraction*" (8, emphasis in text) relative to lower-level properties and laws:

> [M]ultiple realizability is something of a red herring. What matters, fundamentally, is not whether there could be minds embodied in things other than brains, but rather whether there is a level of reality beyond the level at which brains are normally studied—whether psychological kinds are "really there" [in addition to] the already recognized kinds in chemistry, biology and the other established sciences. If this is what is at stake, then it would not matter if brains turned out to be the only kinds of things that realize minds in any nomologically possible worlds. [. . .] The functional descriptions, and the generalizations given in terms of the psychological categories defined at the functional level would still, in this case, be autonomous from the descriptions, generalizations and categories that turned up at the level of the realizers. (8)

This line of thought seems right to me. Taking it on board, one might also say that while multiple realizability is a good indicator of when a comparatively abstract ontological and causal joint is in place, that there is such a comparatively abstract joint does not hinge on multiple realizability. Also worth noting is that causal joints may overlap—both in respect of a given token power and in respect of an

[34] Note that nothing in this line of thought requires that one accept a 'difference-making' or counterfactual account of causation, or relatedly, that one reject *P* as being a cause of *E*. The suggestion is simply that attention to difference-making considerations provides a principled basis for *S*'s being distinctively efficacious with respect to *P*. I'll return to this issue in Ch. 3.

associated effect *E*. If the joints as a whole are different, this provides a principled reason for taking a realized feature *S* to be distinctively efficacious vis-à-vis *E*, in that *S* produces *E* as part of a different system of laws (different causal joint) than *S*'s realizer, *P*. Relatedly, insofar as causal relations involve exercises of a given power, this appears to be what Macdonald and Macdonald (1995) have in mind when they say that the causal autonomy of realized mental features reflects that "any instance of a cause-effect relation can be an instance of more than one pattern" (71).

The key suggestion operative in the nonreductive physicalist's approach is that *there are two ways for a higher-level feature to be distinctively efficacious* with respect to the lower-level feature(s) upon which it depends. One way, emphasized by Kim and others, is for the higher-level feature to be associated with a new power to produce the effect; here the distinctive efficacy (characteristic of accounts of Strong emergence) is located in the having of a distinctive (fundamentally novel) power. Another way—that at issue in the powers-based subset strategy, and characteristic of accounts of nonreductive physicalism/Weak emergence—is for the higher-level feature to be associated with a *distinctive subset of powers* that are relevantly proportional to the effect, in the ways indicated by difference-making considerations and comparatively abstract special-science laws or (more generally) causal joints. In short: distinctive efficacy may reflect either the having of a distinctive *power* or the having of a distinctive *power profile*, and it is the latter variety of efficacy that is at issue in accounts of physical realization.

The nonreductive physicalist response to the problem of higher-level causation

Let's now see in more detail how satisfaction of the *Proper Subset of Powers Condition* enters into the nonreductive physicalist's response to the problem of higher-level causation, and in particular into their rejection of *Non-overdetermination* (according to which, with the exception of double-rock-throw cases, effects are not causally overdetermined by distinct individually sufficient cotemporal causes).

In case 1, special-science feature *S* again depends on base feature *P*, and *S* causes another special-science feature *S**, which depends on base feature *P**; as previously, we might suppose that *S* is a state of feeling thirsty, and *S** is a desire to reach for a nearby glass of water. Here the nonreductive physicalist's strategy involves, to start, the supposition that *S* satisfies the *Proper Subset of Powers Condition* in such a way as to have the power, on a given occasion, to bring about *S**. As per the condition, this token power is identical to one had by *P*; hence when *S* causes *S**, so too does *P*. *S* and *P* are each sufficient causes of *S**; they are distinct, by Leibniz's Law, since *S*, in satisfying the *Proper Subset of Powers Condition*, has fewer token powers than *P*; and since *S* and *P* are cotemporal, they are not parts of a diachronic causal chain. Consequently, *S** is overdetermined, contra *Non-overdetermination*. Yet, the nonreductive physicalist maintains, the overdetermination here is not of the double-rock-throw variety that is supposed to be problematic (since implausible).

In a double-rock-throw case, distinct token powers and associated causal chains converge on a single effect (a window breaking). In cases where the higher-level cause satisfies the *Proper Subset of Powers Condition*, however, two distinct token features standing in an intimate dependence relation are each associated with the same token power and hence the same causing. The overdetermination here is benign—indeed, is no more problematic than in cases where, e.g., both a plane and its wheels are causes of a runway's being touched.[35] Hence it is that *Non-overdetermination*, at least in full generality, must be rejected, leaving the way clear for higher-level efficacy. Here it's worth representing the features at issue as having overlapping sets of powers, with each power represented as a dot, as in Figure 2.5:

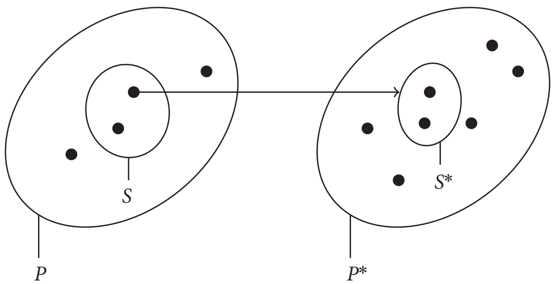

Figure 2.5 The nonreductive physicalist's response to case 1

In case 2, special-science feature *S* rather causes a base feature *P**. In the first instance, the treatment of this case, shown in Figure 2.6, is a variation on the same theme:

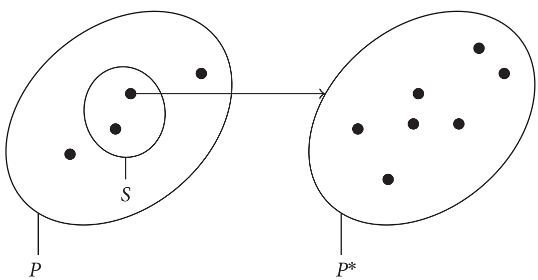

Figure 2.6 The nonreductive physicalist's response to case 2, version 1

[35] Note that, contra a guiding supposition of Morris 2011, the motivations for thinking that satisfaction of the *Proper Subset of Powers Condition* blocks problematic overdetermination do not hinge on this sort of illustrative analogy to cases of benign part/whole overdetermination. Rather, that the overdetermination is benign follows just from the fact that, if the condition is satisfied, only one token power is manifested on the occasion in question, in which case the overdetermination here is nothing like that at issue in double-rock-throw cases.

That said, there is a subtlety here that must be addressed, associated with the possibility that no other lower-level feature besides *P* can, in the relevant circumstances, cause *P**. In this case, the supposed distinctive autonomy of *S* cannot rely on difference-making considerations, according to which if *S* had been realized by some lower-level physical property other than *P*, *S* would still have caused *P**. Nor can *S*'s distinctive efficacy vis-à-vis *P** be a matter of *S* and *P** mutually occupying a distinctively abstract system of laws or causal joint, since *P** is, by assumption, a lower-level physical feature.

In such a case, we can still accommodate the seeming efficacy of *S* vis-à-vis physical goings-on by taking these appearances to concern, not lower-level goings-on (at, e.g., the quantum level), but rather what we might call 'physically acceptable' goings-on *P*′ at some level lower than that at which *S* is properly located, but higher than that at which *S*'s ultimate lower-level physical realizer *P* is located. For example, the nonreductive physicalist can accommodate mental state *S*'s being distinctively efficacious vis-à-vis some sort of physical behavior *P*′—say, a reaching for a glass—which is also realized by *P**, but for which difference-making considerations vis-à-vis *S* would be present. In this case, a more accurate picture would be as shown in Figure 2.7:

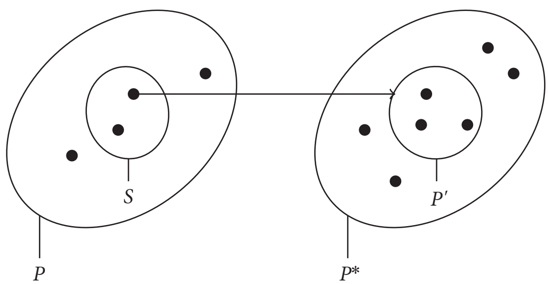

Figure 2.7 The nonreductive physicalist's response to case 2, version 2

Indeed, irrespective of whether *P* is the only lower-level physical feature capable of causing *P**, this sort of model might more generally make better sense of the seeming efficacy of special-science features vis-à-vis physically acceptable goings-on.

It is also worth noting that the strategy here is compatible with physicalism. The main concern about *S*'s physical acceptability, given its reality, cotemporal material dependence on, and distinctness from its dependence base feature *P*, turns on the possibility that *S* might be Strongly emergent from *P*—that is, that *S* might have, as per the *New Power Condition*, a (fundamentally novel) power not had by *P*, of the sort undermining *Physical Causal Closure*. But *S*'s satisfaction of the *Proper*

Subset of Powers Condition is incompatible with *S*'s satisfaction of the *New Power Condition*, ruling out *S*'s being Strongly emergent.[36]

2.3.2 The schema for Weak emergence

Prima facie, satisfaction of the *Proper Subset of Powers Condition* by a special-science feature *S* which cotemporally materially depends on a base feature *P* provides an appropriate and illuminating basis for avoiding overdetermination while guaranteeing that *S* is both ontologically and causally autonomous with respect to *P*. Satisfaction of this condition guarantees that *S* is ontologically autonomous (distinct) from *P*: since *S* has a proper subset of the token powers of *P*, *S* is distinct from *P*, by Leibniz's Law. It moreover makes room for *S* to be causally autonomous (distinctively efficacious) with respect to *P*, in virtue of *S*'s having a distinctive power profile, tracking difference-making/proportionality considerations or a comparatively abstract system of laws or (more generally, and compatible with the distinctive efficacy of artifacts) causal joint.

We have thus arrived at our second schema for metaphysical emergence:

> *Weak Emergence*: What it is for token feature *S* to be Weakly metaphysically emergent from token feature *P* on a given occasion is for it to be the case, on that occasion, (i) that *S* cotemporally materially depends on *P*, and (ii) that *S* has a non-empty proper subset of the token powers had by *P*.

Here again, the locution 'what it is for' is intended to flag that *Weak Emergence* provides a schematic metaphysical basis for a given case of such emergence, encoding what is core and crucial to that notion. Of course, any schematic account needs to be sensibly filled in. Modulo this caveat, the conditions in the schema are, I claim, necessary and sufficient for (appropriate and illuminating accommodation of) metaphysical emergence of the physically acceptable variety.[37] Some clarifications:

[36] As discussed in Wilson 1999, satisfaction of the *Proper Subset of Powers Condition* also appears to block other live routes to physical unacceptability, associated with *S*'s being non-natural (see Moore 1903) or supernatural (as per Malebranchean occasionalism). As it happens, Moore used the term 'non-natural' as indicative of epistemological irreducibility (more specifically: indefinability), which is arguably compatible with physicalism (see Wilson 2002*a*). But supposing that *S*'s being epistemologically irreducible is deemed naturalistically (hence physically) problematic, this must be because such irreducibility indicates that *S*'s existence involves something metaphysically new relative to ('over and above') *P*; on the nonreductive physicalist's operative assumption that *S* is efficacious, the problematic addition in question would presumably either be or entail *S*'s having of a non-natural or supernatural causal power, not had by *P*. But if the *Proper Subset of Powers Condition* is satisfied, the having of such a power is ruled out.

[37] Here again, this is not to presuppose that all implementations of *Weak Emergence* are on a par, or equally well suited to accommodate a given macro-phenomenon. For example, Bealer (1997) argues that functional realization cannot properly accommodate self-consciousness, and Ehring (1996) and others argue that certain features of the determinable/determinate relation render it unsuited as an account of the realization of mental states (though see my responses in Ch. 7, §7.2.3). And as we'll see in Ch. 3, certain implementations of *Weak Emergence* have distinctive advantages.

- Again, the first condition minimally specifies cotemporal material dependence, understood as involving both (physical) substance monism and the minimal nomological supervenience of emergent feature types on base feature types.
- The second condition (effectively, the *Proper Subset of Powers Condition*) captures the comparatively weak sense in which an emergent feature may be both ontologically and causally autonomous with respect to its base feature, in virtue of having a proper subset of the token powers of the base feature.

  Note that while the initial presentation of the *Proper Subset of Powers Condition* built in the cotemporal material dependence of the feature having the proper subset of powers on its base feature, the schema for *Weak Emergence* treats the proper subset of powers and dependence conditions separately. Henceforth, references to the *Proper Subset of Powers Condition* should be understood as references just to condition (ii) in that schema.
- Again, the base feature at issue is a feature of a given (ultimately physical micro-) configuration (plurality or structural aggregate), and the overlap in powers at issue in condition (ii) should be understood accordingly. In particular, this overlap in powers is not to be (implausibly) understood as involving overlap of powers of a macro-entity with powers of features of individual entities entering into the configuration upon which the macro-entity cotemporally materially depends.
- Again, the schema is relativized to occasions (times or temporal intervals), but it is reasonable to suppose that (given that *S*'s type is not Strongly emergent) it suffices for the Weak emergence of *S*, *simpliciter*, that the condition is ever satisfied, and to suppose that it suffices for the Weak emergence of the feature type (of which *S* is a token), *simpliciter*, that any token feature *S* on any occasion satisfies (or would satisfy) the condition. These complications won't play a role in what follows.

2.4 The schemas for Strong and Weak emergence as core and crucial to metaphysical emergence

Let's sum up the results so far. Attention to the problem of higher-level causation points towards two strategies of response to this problem, associated with Strong emergentism and with nonreductive physicalism, respectively, each of which provides an appropriate and illuminating basis for accommodating the metaphysical emergence of higher-level entities and features, understood as coupling cotemporal material dependence with ontological and causal autonomy (that is, with distinctness and distinctive efficacy). Each response is associated with a specific condition on the token powers of a higher-level feature with respect to the token powers of the lower-level feature upon which it depends on a given occasion, which

condition, along with a minimally specified condition on cotemporal material dependence, is encoded in the associated schema for Strong or Weak emergence:

> *Strong Emergence*: What it is for token feature *S* to be Strongly metaphysically emergent from token feature *P* on a given occasion is for it to be the case, on that occasion, (i) that *S* cotemporally materially depends on *P*, and (ii) that *S* has at least one token power not identical with any token power of *P*.
>
> *Weak Emergence*: What it is for token feature *S* to be Weakly metaphysically emergent from token feature *P* on a given occasion is for it to be the case, on that occasion, (i) that *S* cotemporally materially depends on *P*, and (ii) that *S* has a non-empty proper subset of the token powers had by *P*.

Indeed, attention to these responses makes clear the limited ways in which a cotemporally materially dependent higher-level feature can be causally, hence ontologically, autonomous with respect to its base feature, as the operative conception of metaphysical emergence requires. First, the feature may have *more* powers than its base feature; second, the feature may have *fewer* powers than its base feature. In terms of effects: the higher-level feature may be distinctively efficacious in potentially contributing to causing *more* effects than its base feature, or it may be distinctively efficacious in potentially contributing to *fewer* effects than its base feature. Since complete coincidence of token powers doesn't make room for causal autonomy (distinctive efficacy), these routes to metaphysical emergence exhaust the available options.

I conclude that satisfaction of the conditions in either schema is, as I put it, 'core and crucial' to metaphysical emergence of the sort relevant to realistically vindicating the seeming appearances of emergence as pertaining to special-scientific and artifactual entities and features. Modulo the supposition that the schemas are sensibly filled in, the results of this chapter can be seen as providing prima facie reason to think that the conditions in the schemas are both necessary and sufficient for (appropriate and illuminating accommodation of) metaphysical emergence of both physically acceptable and physically unacceptable varieties—a bold claim, but one that, as I argue in ensuing chapters, is surprisingly robust.

We thus have a provisional schematic answer—or rather, two answers—to the first key question, 'What is metaphysical emergence'? The answers are schematic, since as we have already seen and will see further in the chapters to follow, there are several ways in which either schema might be (or might aim to be) implemented. The answers are also at this point provisional, since a number of objections have been raised to the viability of physically acceptable or physically unacceptable emergence, either in general or as specifically directed at the schemas for *Weak* or *Strong* emergence or the conditions on powers therein. In the next two chapters, I'll consider these objections, and show that they can be answered.

3

The viability of Weak emergence

In Ch. 2, I provided prima facie reasons for thinking that satisfaction of the conditions in the schema for Weak emergence is core and crucial to a feature's being metaphysically emergent from—cotemporally materially dependent on, yet ontologically and causally autonomous with respect to—a lower-level feature, in a way compatible with physicalism, given that the base feature is physically acceptable. Again, the schema is as follows:

> *Weak Emergence*: What it is for token feature *S* to be Weakly metaphysically emergent from token feature *P* on a given occasion is for it to be the case, on that occasion, (i) that *S* cotemporally materially depends on *P*, and (ii) that *S* has a non-empty proper subset of the token powers had by *P*.

Henceforth, by 'Weak emergence' I will intend to refer to the view (or the phenomenon, understood in light of the view) that conformity to the conditions in *Weak Emergence* is core and crucial—and when sensibly filled-in, both necessary and sufficient—for physically acceptable emergence (i.e., realization).

In this chapter, I consider and respond to a representative range of objections that have been or could be made to the viability of Weak emergence, so understood. These objections fall into four main categories, according to which satisfaction of the conditions in *Weak Emergence* is compatible with anti-realism about higher-level features (§3.1), is compatible with reductionism about higher-level features (§3.2), is compatible with the emergent feature's being physically unacceptable (§3.3), or is not necessary for metaphysical emergence of a physically acceptable variety (§3.4). The primary focus of many of the objections is on condition (ii) in the schema—i.e., the *Proper Subset of Powers Condition*. As we'll see, each of these objections admits of at least one response that could be endorsed by any proponent of Weak emergence, whatever their preferred implementation of the schema. Upon occasion, however, a response to a given objection is available that relies on a specific implementation of the schema. In particular, certain attractive responses appeal to either a determinable-based account of Weak emergence or a degree of freedom-based (DOF-based) account of such emergence, of the sorts I have previously endorsed.

Metaphysical Emergence. Jessica M. Wilson, Oxford University Press (2021). © Jessica M. Wilson.
DOI: 10.1093/oso/9780198823742.003.0003

3.1 Objection: compatibility with anti-realism

Certain philosophers, who seem to me to be properly deemed anti-realists, maintain that seeming satisfaction of the *Proper Subset of Powers Condition* reflects not an independently existing higher-level feature, but rather a (mere) mind-dependent abstraction from, or pragmatically motivated way of conceiving of, lower-level physical goings-on.

For example, Ney (2010) argues, after discussing how Shoemaker (2000/2001, 2003, and 2007) takes satisfaction of the *Proper Subset of Powers Condition* to serve as the basis for realization, that an alternative way of understanding seeming satisfaction of this condition is available:

> [One] can [...] think of things in the following way (nothing has been said to rule out this way of thinking): on this view, really the mental event (and realized tokens more generally) are just abstractions from concrete microphysical situations. They are abstractions in the sense that they are what we attend to when we focus only on a proper subset of a microphysical state's causal powers. (442)

Ney sees this reading as indirectly supported by Shoemaker's (2007) claim that realization relations form a hierarchy, where "those higher in such a hierarchy will be realized by those further down" (23), and where properties whose instantiations lie at the bottom of the hierarchy are, unlike other properties, "self-constituted". As she says:

> Plausibly, [one] can tolerate the existence of entities that are not reducible, i.e., identical to, physical entities, so long as these things are mere abstractions. Since realized events can very naturally be seen on Shoemaker's account as mere abstractions from their realizer events, since they are not self-constituted, I don't see why [one] cannot endorse this approach. (443)

On such an interpretation, "The only entities with genuine, (mind-)independent existence here are the microphysical states of affairs" (444). Though there might be a sense in which, on Ney's interpretation, mental states understood as 'mere' mind-dependent abstractions exist, in the sense relevant to our investigations here this interpretation is properly deemed 'anti-realist'.[1]

[1] Ney herself presents such abstractionism as compatible with reductionism, but in my view this muddies the terminological waters, insofar as standardly, and as I am understanding it here (and as Ney herself suggests in the previous quotation), ontological reductionism entails that the reduced goings-on are identical to the reducing goings-on. Similar remarks apply to Heil's view (to be next discussed), which he sometimes presents as being compatible with a form of realism about higher-level features.

Heil (2003*b*) also argues for a kind of anti-realist conceptualism about higher-level features. As Heil sees it, embrace of nonreductionism reflects an uncritical tendency for philosophers and others to suppose that "we can 'read off' features of reality from our ways of speaking about it" (207), in a way that involves the following 'Picture Theory', whereby representation always corresponds to reality:

> *The Picture Theory*: When a predicate applies truly to an object, it does so in virtue of designating a property possessed by that object and by every object to which the predicate truly applies (or would apply). (210)

By way of illustration of what is problematic about this principle, Heil argues that notwithstanding that the predicate 'is red' truly applies to some objects—e.g., tomatoes, stoplights, apples—it would be a mistake to uncritically suppose that this is in virtue of these objects' possessing the "*very same* property". Rather, he suggests, one may take the predicate and associated concept to here be serving the broadly pragmatic function of enabling us to categorize objects that are inexactly similar:[2]

> The concept expressed by the predicate 'is red' [...] seems tailor-made for picking out a range of objects that are, in a particular way, less-than-perfectly-similar to one another. The concept applies to objects by virtue of properties possessed by those objects, presumably an extremely complex and diverse class of physical properties. There is, I gather, no prospect of defining or analyzing redness in terms of these physical properties. This is due, in some measure, to the fact that the properties in question are salient—to us—partly owing to the nature of our perceptual system. Were we built differently, were we made of different materials, the diverse collection of properties that satisfy our concept of redness could well fail to stand out. In that case we should have no use for the concept. (215)

The same holds more generally, Heil continues, for special-science predicates such as 'is a tree', 'is a planet', or 'is in pain'. Rather than taking diverse applications of a given such predicate as indicating the existence of a common property—whether reducible or irreducible, no matter—we should see such applications as reflecting a merely conceptual means of tracking inexact similarity among lower-level physical properties.

[2] Berkeley (1710) argues similarly against the assumption that general terms denote abstract ideas: "[I]t is thought that every name has, or ought to have, only one precise and settled signification, which inclines men to think there are certain *abstract*, *determinate* ideas, which constitute the true and only immediate signification of each general name [...] whereas, in truth, there is no such thing as one precise and definite signification annexed to any general name, they all signifying indifferently a great number of particular ideas" (§18).

Heil here endorses (what is reasonably described as) a form of anti-realism about higher-level features, at least in cases of seeming multiple realizability. Such cases are frequently appealed to in support of the holding of the *Proper Subset of Powers Condition* (with the powers associated with the higher-level feature being those in the intersection of the sets associated with its multiple realizers); hence Heil's deflationary conceptualism constitutes an objection to the viability of Weak emergence, to the effect that insofar as the condition on powers appears to be satisfied, this has at best pragmatic implications.

My response to the anti-realist begins by first granting that, as Ney puts it, "nothing has been said to rule out" an abstractionist or pragmatist line on seeming satisfaction of the *Proper Subset of Powers Condition*; but second, denying that the viability of Weak emergence hinges on accomplishing such a 'ruling out'. As per the operative methodological criterion of appropriate accommodation (Ch. 1, §1.4.5), I take the natural (straightforward, default) accommodation of the appearances of metaphysical emergence to be a heavily weighted desideratum—more weighty, in particular, than any antecedent parsimony-based motivation, according to which one should strive to minimize the number of ontological posits invoked in a metaphysical account of some phenomenon, independent of considerations of the usual understanding of the phenomenon, and in the absence of any identified problems with such posits. Correspondingly, to the extent that the appearances support the posit of metaphysically emergent features, the burden is on the anti-realist to provide reasons for not taking the appearances at face value.[3] As I'll now argue, however, neither Ney nor Heil provides any such good reason—in particular, as telling against a Weak emergentist treatment of the appearances.[4]

To start, one might wonder whether Heil's 'debunking' explanation of motivations for nonreduction, according to which these motivations really reflect an unreflective inclination to take any true predication as tracking a genuine property, would count as undermining the operative presumption of the criterion of appropriate accommodation. By way of brushclearing, it is worth noting that the focus on predicates is not really to the point, since predicates or concepts do not after

[3] This is even more true in the present case, where the considerations offering prima facie support for there being metaphysical emergence are very extensive, drawn from the special sciences, perception, linguistics, and ordinary practices of individuation, among others. I doubt that all these phenomena would admit of anti-realist treatment, but even if some such treatments could be cooked up, they would be to some extent revisionary. For example, on Heil's approach, special-science laws which appear to make no reference to lower-level entities and features would turn out to really be about such entities and features (and their inexact similarities, about which more shortly).

[4] Morris (2018) offers a developed 'argument by cases' for the conclusion that physicalism must take the form of a 'one-level' anti-realist (or reductionist; as in Ney and Heil's discussions, the distinction here is somewhat fluid) view similar to that endorsed by Ney and Heil, which proceeds by raising concerns for a range of accounts of realization, including what Morris calls 'subset realization'. In this section I'll mention two salient concerns raised by Morris (one drawn from his 2013); what I take to be his most pressing concerns, however, are best treated in §3.2.3 and §3.3.1.

all appear from nowhere, to be interpreted (or misinterpreted, as the case may be) as referring to genuine features, but are rather typically introduced into scientific and ordinary discourse in response to a felt need to refer to features that we have some or other independent reason to think exist. Heil's concern, more properly understood, is that any seeming motivations for positing higher-level or more specifically metaphysically emergent entities or features are better interpreted simply as tracking inexact similarities between lower-level entities or features.

This 'debunking' account of the seeming motivations for metaphysical emergence is unsuccessful, however. For even granting, e.g., that seeming attributions of higher-level features were in some sense tracking inexact similarities among lower-level features, this would not show that the higher-level features did not exist, unless it was antecedently clear that the inexact similarities at issue were not themselves higher-level. But this is not antecedently clear; on the contrary, someone aiming to take the appearances at realistic face value might happily grant that higher-level features are in some sense tracking inexact similarities between lower-level features, but maintain that such inexact similarities are themselves metaphysically emergent features of reality. Upon closer examination, then, neither considerations pertaining to the Picture Theory nor considerations pertaining to the connection between higher-level feature attributions and inexact lower-level similarities provide independent reason to reject the prima facie appearances of higher-level reality, beyond the background methodological supposition (which we are here rejecting) according to which ontological parsimony is weighted more heavily than a natural accommodation of the appearances.

Heil's anti-realism is also motivated by Kim-style concerns that higher-level properties are (on pain of problematic causal overdetermination) causally excluded by their lower-level realizers:

> I need not remind you of difficulties a levels conception breeds. Consider just the problem of the 'causal relevance' of higher-level properties. Suppose that mental properties are higher-level properties realized in the nervous systems of sentient creatures. How could such properties affect the behavior of creatures possessing them? The potential causal contribution of any higher-level property would seem to be preempted by its lower-level realizing property. [...] This worry about the causal relevance of mental properties extends smoothly to higher-level properties generally. If you like to think of the special sciences as occupied with higher-level properties and events, then you will need some accounting of how these properties and events could make a causal difference in our world.
>
> (Heil 2003*b*, 213)

In Heil's remarks, we again see operative the assumption, following Kim, that the only way for a feature to be distinctively efficacious is by its having a distinctive

power. But as previously, the Weak emergentist has provided an "accounting of how [higher-level] properties and events could make a causal difference in our world", in spite of not having any distinctive powers—namely, by having distinctive power profiles, reflecting either causal difference-making considerations and/or comparatively abstract, metaphysically real, systems of laws or causal joints. What is needed at this point, then, is anti-realist reason to think that the strategy of locating distinctive efficacy in the having of distinctive power profiles is somehow problematic.

Ney offers such a reason, but in focusing solely on Shoemaker's discussion, her objection misses the mark. After presenting her mental abstractionist alternative, she says, on Shoemaker's behalf:

> But maybe this is too quick. At times Shoemaker has tended to suggest that in cases of mental causation, it is only the mental event [*S*] that is efficacious with respect to the effect. In *Physical Realization* [Shoemaker 2007], he makes the slightly weaker but similar suggestion that in cases of mental causation, only the mental event is directly efficacious, with the physical or microphysical realizer events being efficacious only in virtue of containing the causal powers of the mental event as parts (53). My speculation is that the reason Shoemaker makes claims like these is in order to emphasize that he is really providing a nonreductive account of mental causation [...] The goal is to do this by securing some kind of distinctive causal efficacy for the mental, claiming that in cases of mental causation, it is only mental events (not underlying physical or microphysical events) that are efficacious, or that in cases of mental causation, it is only mental events that are directly efficacious.[5] (443)

Ney goes on to say that, insofar as every power of a mental feature on a given occasion is supposed to be identical with a power of its physical realizer, and insofar as satisfaction of this token identity condition is crucial to preventing overdetermination, "it doesn't make sense to say that the realizers are only indirectly efficacious vis-à-vis the efficacy of the mental event. There is only one causal relation here [...] there is no reason to try to secure any distinctive causal efficacy for mental events. Assuming we are all physicalists, the challenge is to secure nonredundant causal *efficacy* for mental events, not causal distinctiveness" (444). Hence, she concludes, Shoemaker's strategy for securing the distinctive efficacy of higher-level features, via satisfaction of the *Proper Subset of Powers Condition*, fails.

[5] Morris (2011) similarly suggests "perhaps we could take Shoemaker to be saying that [...] when a property *M* is subset realized by a property *P* on some occasion, and *M* and *P* seem to overdetermine an effect, in fact it is the *M*-instance that is the 'real' cause of that effect" (370).

It is true that Shoemaker has several times suggested that a higher-level feature satisfying the *Proper Subset of Powers Condition* might be appropriately deemed 'the' cause of a given effect. For example, in his (2000/2001), Shoemaker says

> [W]here only the causal features of a property *P* that play a role in producing an effect are ones that belong to a property *M*, of which *P* is a [...] realizer property, there seems a good sense in which considerations of proportionality [including sufficiency and difference-making] favor the instantiation of *M* over the instantiation of *P* as the cause of the effect. (436)

In his (2003), he confirms:

> One advantage of this approach is that it provides a basis for saying that in some cases it is the instantiation of a mental property, rather than the instantiation of one of its realizer properties, that caused a certain effect, or contributed to causing it. (3)

But here, in my view, Shoemaker's discussion goes awry. I agree with Ney that the supposition that a Weakly emergent feature might be appropriately deemed 'the' cause (or a 'more direct' cause) of a given effect is problematic: she is right that it doesn't make sense to deny that the lower-level realizer is a cause (and moreover a direct cause) of the effect, given that the supposition that the *Proper Subset of Powers Condition* is satisfied entails that the realizer has the power to produce (or contribute to producing) the effect, and this power is manifested (or the associated regularity instanced, etc.) on the occasion in question. Moreover, as I note in Wilson 1999, 2011*b*, and elsewhere, the supposition that a higher-level feature causes something that its base feature does not cause is, for reasons that Kim's causal overdetermination concern brings out, in tension with physicalism and the associated supposition of *Physical Causal Closure*, whether the purported effect is a higher-level special-science feature or a lower-level physically acceptable feature.

There is, however, no need to follow Shoemaker in suggesting that realized features can cause effects that their realizing features don't cause, since what the Weak emergentist requires is not that higher-level features be *uniquely* efficacious but just that they be *distinctively* efficacious. For similar reasons, the Weak emergentist can and should reject Yablo's (1992) claim that difference-making and other proportionality considerations can support identifying a determinable feature (e.g., red) rather than an associated determinate feature (e.g., scarlet) as *the* cause of a given effect (e.g., the pecking of a pigeon at a scarlet patch), as well as List and Menzies' (2009) account of higher-level autonomy as involving 'downwards exclusion':

> If it is correct that realization-insensitivity is a general requirement in higher-level causal claims, then it follows that the conditions for downwards exclusion are generally satisfied. But this in turn entails that higher-level causal relations such as that between [realized] *M* and [effect] *B* obtain even though there is no underlying causal relation between the neural realizer *N* and *B*. In such cases, we have good reason to believe in the causal autonomy of higher-level properties.[6]
> (499)

Once it is seen that distinctive efficacy can be achieved either via the exercise of a distinctive power *or* via a distinctive power profile (and associated difference-making and/or abstract causal joint/law considerations), one can allow that both realizing and realized features are causes—distinctive causes—of a given effect: one can have one's physicalism and one's distinctive higher-level efficacy, too.[7]

Having gotten clear about the strategy operative in Weak emergence for achieving distinctive efficacy of higher-level features, one can also see that Ney's discussion fails to accurately take this strategy into account. This distinctive efficacy is not (even for Shoemaker, notwithstanding certain of his remarks' suggesting otherwise) a matter of a higher-level feature's being able to cause things that its realizer(s) cannot cause; rather, it is a matter of its profile's being such as to track relevant difference-making considerations, or comparatively abstract systems of laws or causal joints. Relatedly, Ney is wrong to claim that for physicalists "the challenge is to secure nonredundant causal *efficacy* for mental events, not causal distinctiveness", for ensuring the former without ensuring the latter will still fail to accommodate the prima facie appearances of emergence, contra our guiding methodology, and relatedly, will lead to the need to drastically revise our under-

[6] It is worth noting that even if one endorses a response to Kim-style overdetermination concerns appealing not just to difference-making considerations (as per, e.g., Horgan 1989 and Lepore and Loewer 1987 and 1989; see also Bennett 2003) but moreover to a difference-making or counterfactual account of causation (as List and Menzies do), one might not be committed to gaining the distinctive efficacy of a higher-level feature vis-à-vis a given effect at the cost of denying that its physical dependence base feature is also a cause of the effect (contra *Physical Causal Closure*); for one may rather suppose that higher-level and lower-level features are each causes, relative to different contrast classes (see, e.g., Hitchcock 1996.)

[7] My response to Ney applies, mutatis mutandis, to the 'symmetry' concern raised in Morris (2013), according to which attention to causal implications of cases of the determinable/determinate relation fails to indicate that higher-level features, on a determinable-based account of realization, will satisfy the *Proper Subset Condition on Powers*. Morris correctly observes that if (as both Yablo 1992 and Shoemaker 2000/2001 suggest, by appeal to proportionality considerations), e.g., the patch's being red is the best candidate for being the cause of Sophie's pecking and the patch's being scarlet is the best candidate for being the cause of Alice's pecking, then we have symmetry here, with each feature having at least one power that the other doesn't have, contra the supposed satisfaction of the *Proper Subset of Powers Condition*. Like Ney, Morris assumes that in the cases at hand, the assignment of a power to an instance of a feature depends on taking the feature to be *the* cause of the effect in question: *red* but not *scarlet* causes Sophie to peck, while *scarlet* but not *red* causes Alice to peck. Morris is correct that this reading of the cases is problematic, but as per the discussion in the main text this reading can and should be rejected.

standing of the truth and/or content of a large range of scientific and ordinary beliefs.

Unlike Ney, Morris (2018) registers the strategy of appealing to difference-making considerations as a means of gaining the distinctive efficacy of subset-realized features in a way compatible with nonreductive physicalism/Weak emergentism, but maintains that more work needs to be done to show that such a strategy "can provide a nonreductive response to the exclusion problem" (154), as involving mental features, in particular:

> [C]an the powers of a mental occurrence really be a subset of the powers of a physical occurrence, if the former is a difference-making cause of certain events while the latter is not? Even if an affirmative answer is possible, talk of mental and physical occurrences only involving a single "causing" becomes strained if mental causes are difference-making causes while physical causes are not—how could a mental occurrence's bringing about some event involve the very same "causing" as some physical occurrence's bringing about that event, if the one is more proportional or difference-making than the other with respect to that event? (154, note 32)

In response to Morris's initial question, I say 'yes': satisfaction of the *Proper Subset of Powers Condition* is not only compatible with cases of higher-level difference-making, but in the motivating cases—that is, cases where a mental or other higher-level feature is properly seen as tracking certain difference-making considerations in the production of a given effect that both it and its lower-level realizer have the power to cause (and manifest, on the occasion in question)—provides the basis for explaining why the higher-level feature is distinctively efficacious in this way. It is because the higher-level feature has fewer powers than its realizer (on the occasion in question, and so on), such that changes in powers not shared between realizer and realized don't matter—don't make a difference—to the production of the effect, that the higher-level feature tracks difference-making considerations. Nor does it seem to me that when a higher-level feature is distinctively efficacious in this way, there is any difficulty in understanding how both the higher- and lower-level features could cause the very same effect. The having of a power is one thing—one way for a feature to be efficacious; the having of a power profile is another thing—another way for a feature to be efficacious. Correspondingly, two properties can exercise the same power on a given occasion, but only one have a power profile tracking difference-making considerations (or comparatively abstract systems of laws or causal joints), as cases such as that of the differently trained pecking pigeons illustrate. If there is more work to be done here, it lies in identifying some specific problem for this means of accommodating the causal autonomy at issue in physically acceptable emergence.

So far, then, we have not been given good reason to resist taking the prima facie appearances of higher-level reality seriously, as treated along lines of Weak emergence, in particular.[8]

3.2 Objection: compatibility with reductionism

In this section, I consider objections targeting Weak emergence on grounds that even granting that satisfaction of the conditions in *Weak Emergence* by features *S* and *P* is realistically understood, such satisfaction is compatible with *S*'s being ontologically reducible to—that is, identical with—some *other* lower-level physically acceptable feature *P*′. The general line of thought is as follows. Let the level of base property *P* be the fundamental physical level—the only level the reductionist thinks exists. To be sure, if *S* is Weakly emergent from *P* on a given occasion, then *S* has, on that occasion, a proper subset of the token powers of *P*. Still, the reductionist maintains, given that *S* satisfies the *Proper Subset of Powers Condition*, surely *S* must be reducible to—that is, identical to—some *other* physically acceptable feature *P*′. As Yates (2012) puts the concern:

> [I]f *P* has as a constituent a physical property *P*′ whose powers are the same proper subset of *P*'s that *M* inherits, then there's nothing to prevent identification of *M* with *P*′. (6)

Whether both *P* and *P*′ should be considered 'realizers' of *S* is a topic for further debate.

In what follows I present and address three different strategies for achieving such a reduction.

3.2.1 Reduction to a conjunct of a lower-level conjunction?

On the first reductive strategy, satisfaction of the *Proper Subset of Powers Condition* is compatible with seemingly higher-level feature *S*'s being identical with a

[8] Here I put aside another motivation for anti-realism about (some) higher-level features, associated with eliminativist physicalism, according to which if we cannot explain a seemingly existent higher-level feature in physical terms—if an insuperable explanatory gap exists between the seeming feature (or associated laws and theories) and lower-level physical features (laws, theories)—then we should reject the seeming existence of the higher-level feature as genuine (see, e.g., Feyerabend 1963, Paul Churchland 1981, and Patricia Churchland 1986). Insofar as the primary eliminativist target has been qualitative mental features, this objection is not generally directed against the viability of Weak emergence, since many special-science features, of the sort that would be candidates for satisfying the schema for Weak emergence, are uncontroversially explainable in lower-level physical terms (and as I'll argue in §3.2.3, such explainability, including deducibility, is compatible with Weak emergence). I revisit the question of the metaphysical import of explanatory gaps in Ch. 7 (§7.1).

conjunct P' of a lower-level conjunctive feature P (of form $P'(= S) \wedge P''$), such that S, though different from P, would nonetheless be identical to a lower-level physical feature. Indeed, one might reasonably suppose that conjunct features typically satisfy the *Proper Subset of Powers Condition* with respect to associated conjunctions, insofar as conjunctive features, being more specific, can contribute to producing more effects (in the same circumstances) than their conjunct features can individually do, for reasons similar to those motivating taking (more specific) determinates to have more powers than their associated (less specific) determinables. So, for example, the conjunctive feature *being massy and being charged* has powers to produce effects that *being massy* alone does not have (e.g., attract pins) and that *being charged* alone does not have (e.g., attract other massy objects). But if P is a conjunctive lower-level physical feature having $P'(= S)$ and P'' as conjuncts, then plausibly, S will also be a lower-level physical feature, and so reducible rather than emergent.[9]

That conjuncts of conjunctions may satisfy the *Proper Subset of Powers Condition* is especially pressing for Shoemaker, who characterizes realization in terms just of satisfaction of the *Proper Subset of Powers Condition*. It is also a difficulty for accounts of Weak emergence involving functional realization, since if the higher-level functional role is associated with a lower-level conjunct feature, one might naturally suppose that any conjunction containing that conjunct will count as implementing or realizing the role.

Shoemaker (2000/2001) is sensitive to this concern, and in response stipulates that for a feature to count as realized, it must not be a conjunct of a conjunctive realizer:

> Property P realizes property S just in case the conditional powers bestowed by S are a subset of the conditional powers bestowed by P (and P is not a conjunctive property having S as a conjunct).[10] (78)

The need to stipulatively rule out conjunct/conjunction cases as genuine cases of Weak emergence—i.e., to rule out conjunctive realization—is perhaps a bullet one could bite. There are, however, two non-stipulative strategies for ruling out such cases.

The first strategy, due to Umut Baysan, is compatible with any implementation of *Weak Emergence*. Here one rules out cases of conjunctive realization as satisfying the conditions in the schema by saying more about the intended notion of cotemporal material dependence operative in condition (i). However exactly such dependence is cashed out, in a context where physicalism is at issue the operative

[9] Note that the concern here remains even if, as Gibb (2013) compellingly argues, it is not always the case that conjunctive features have more token powers than each of their conjunct features.

[10] Here I have replaced Shoemaker's variables with S and P for consistency with my discussion.

understanding should be one according to which cotemporally materially dependent features are less fundamental than their base features. But if a conjunctive feature exists at a given level, it is plausible that the conjuncts of that feature are as fundamental as or more fundamental than the conjunctive feature. As such, a conjunct feature *S* would not be appropriately taken to satisfy the relevant condition on dependence in the schema for Weak emergence, notwithstanding that a conjunctive feature (at least nomologically) necessitates its conjuncts; hence such features pose no threat to the sufficiency of this schema, properly interpreted. As Baysan (2014) puts it:

> [A] property *P* realizes a property *Q* if and only if the causal powers of *Q* are a proper subset of the causal powers of *P*, and *P* is more fundamental than *Q*. Thanks to the requirement that a realized property is less fundamental than its realizers, two things that the original version of the subset view cannot explain are guaranteed: first, fundamental properties are not realized; second, arbitrary conjunctions of properties do not realize their conjuncts. (2)

I like the idea of filling in the operative dependence relation at issue, since as it stands, the notion of cotemporal material dependence at issue in condition (i) of *Weak Emergence* is fairly weak. As above, Shoemaker characterizes realization just in terms of the *Proper Subset of Powers Condition*—plausibly because he supposes that this condition also ensures satisfaction of the relevant dependence condition (see also Clapp 2001). But this supposition is dispensable: nothing prevents a proponent of Weak emergence from offering an independent characterization of the cotemporal material dependence at issue, with the *Proper Subset of Powers Condition* serving mainly to ensure the autonomy that is characteristic of emergence. It would, after all, be unsurprising that two conditions would be needed to characterize metaphysical emergence of whatever variety: cotemporal material dependence is one thing, ontological and causal autonomy another.

The second strategy aimed at blocking conjunctive realization is one appealing to a specific implementation of *Weak Emergence* on which conjunct/conjunction cases are non-stipulatively excluded. A determinable-based account of Weak emergence, on which base features are taken to be determinates of higher-level determinables, would work for this purpose, for it is definitive of the determinable/determinate relation that it is not properly metaphysically characterized in terms of anything like the conjunct/conjunction (or relatedly, genus/species) relations (see Wilson 2017 for discussion). In mereological terms: determinates, unlike classical wholes, do not satisfy 'weak supplementation', according to which a whole having one proper part must also have at least one other proper part disjoint from the first (see Simons 1987 for discussion). In Karen Bennett's memorable terms (pers. comm.): determinates are not determinables with frosting on top. As such, on a determinable-based implementation of *Weak Emergence*, the nature of

determinate features alone non-stipulatively rules out that conjuncts of lower-level conjunctions would count as Weakly emergent.

An implementation of *Weak Emergence* in terms of an elimination in degrees of freedom (DOF), as proposed in Wilson 2010*b* and as I will describe in Ch. 5, might also non-stipulatively rule out conjunctive realization, if no features satisfying the condition on degrees of freedom at issue in this account also stand in the conjunct/conjunction relation. Since which degrees of freedom are associated with which features is an empirical matter, the exclusion here would be contingent as opposed to following (as in the case of the determinable/determinate relation) just from features of the specific relation taken to satisfy the *Proper Subset of Powers Condition* itself; but even so the exclusion would not be stipulative.

Summing up: that conjuncts can satisfy the *Proper Subset of Powers Condition* with respect to associated conjunctions poses a prima facie problem for taking satisfaction of the conditions in *Weak Emergence* to be sufficient for metaphysical emergence, since the relation between conjuncts and conjunctions is plausibly one according to which conjuncts are at least as fundamental as their associated (here, presumed physical) conjunctions. The concern can be addressed, however, either by stipulating the exclusion of such cases (a less than optimal but non-fatal addendum), by denying that conjunct features properly satisfy the cotemporal material dependence condition (a reasonable addendum in cases where the dependence at issue is supposed to conform to physicalism), or by endorsing an implementation of *Weak Emergence* (along determinable-based and perhaps also DOF-based lines) that non-stipulatively rules out conjunct/conjunction cases as instances of such emergence.

3.2.2 Reduction to a disjunction of lower-level disjuncts?

The next reductive strategy takes as its starting point an account on which purportedly higher-level features are identical with disjunctions of base features. The strategy is commonly motivated by a seeming objection to reductionism, according to which the multiple realizability of a given type rules out the type's being identical with any one of its realizing types. As Antony (2003) expresses the objection:

> [M]ultiple realizability has *ontological* consequences. Clearly, a property P cannot be identical with a property Q if there can be instances of P that are not instances of Q. But to say that a property S is multiply realizable [by types P_1 and P_2] is to imply that [...] there can be instances of S that are not instances of P_1, and instances of S that are not instances of P_2, and S cannot be identical with either P_1 or P_2. (3)

A popular reductionist response implements the 'disjunctive strategy' or 'disjunctive move' (see Fodor 1974, Jaworski 2002, and Dosanjh 2014 and 2019), according to which multiply realizable types may be identified with a lower-level *disjunctive* type, where each disjunct is a type of lower-level realizer of *S*'s type.[11] As Heil (1992) describes the strategy (without endorsing it):

> Multiple realizability [...] need not deter a determined identity theorist [that is, reductionist]. It is open for such a theorist, for instance, to argue that the relevant [...] characteristic is, in fact, disjunctive in character. That is, it might be that, in you, mental feature *S* is realized in neural structure *N*, whereas in an octopus, *S* is realized in a different sort of neural structure *N*′. Would this undermine type identity? It would not, unless we assume that *S* [could] not be identical with the disjunctive characteristic $\langle N \vee N' \rangle$. (64)

Now, a number of objections have been raised against the disjunctive strategy.[12] Even putting these concerns aside, there is as yet no clear problem for Weak emergence here; for on the usual understanding of what it is for a disjunctive type to be instanced or tokened, the disjunctive strategy is incompatible with satisfaction of the *Proper Subset of Powers Condition*. On the usual understanding, what it is for a disjunctive feature type *D* to be tokened on a given occasion is for one of the disjunct types to be tokened on that occasion. In the case at hand, the disjunct types (P_1, P_2, ..., P_i) correspond to the physically acceptable realizers of *S*'s type. It follows that if *S*'s type is identified with disjunctive type *D*, then any token of *S*'s type would, on a given occasion, be identical with a token of one of its realizer types, on that occasion. But then the *Proper Subset of Powers Condition* would fail to be met: *S*'s token powers would be, on any given occasion, the same as, rather than a proper subset of, the token powers of the base feature realizing *S* on that occasion. But by assumption, *S* satisfies the condition. It follows that on the assumption that *S* satisfies this condition, as Weak emergence requires, *S*'s type cannot be identified with a type consisting of a disjunction of *S*'s lower-level realizers.

[11] See Polger and Shapiro 2016 for arguments that empirical reasons to think that special-scientific types (in particular, mental state types) are multiply realizable are overblown, and that many such cases can be accommodated in type-theoretic terms, one way or another. Discussion of this argumentation is beyond the scope of this investigation; for present purposes I will just note that some of Polger and Shapiro's cases are subject to challenge (see Levin 2016) and that at any rate they are clear that some special-science or artifactual types are multiply realizable by their lights, and that moreover there might well be cases of mental multiple realizability. Correspondingly, I do not treat them here as general proponents of a reductionist approach to higher-level goings-on.

[12] These include that disjunctive features don't exist (as per Armstrong 1978, 19–23), that disjunctive features are not available for purposes of physicalist reduction of scientific properties, because they are too heterogenous to form a natural kind (as per Putnam 1967, Fodor 1974, and Kim 1992*b*), that disjunctions are 'open-ended' (having an indefinite or infinite number of disjuncts), rendering them unsuitable to enter into type-identities (as per Lepore and Loewer 1987 and Pereboom and Kornblith 1991), and that among the metaphysically possible realizers of a multiply realizable feature type such as *S* will be some physically unacceptable types (as discussed, though not endorsed, in Dosanjh 2014).

Granting that the *Proper Subset of Powers Condition* is met, a disjunctivist gambit remains. Drawing in part on discussion in Clapp 2001 and Antony 2003, Dosanjh (2014) argues that it is not generally the case that what it is for a disjunctive property to be instanced is for one of its constituent disjunct properties to be instanced. To be sure, Dosanjh allows, this is the case when the disjuncts of disjunctive properties are gerrymandered or otherwise dissimilar (as with, e.g., the properties *being red or round* and *being sour or prime*). But sometimes, he maintains, the instantiation of the disjunction is *not* the same as the instantiation of one or other disjunct. In particular, when the disjuncts are relevantly similar, in sharing powers associated with the purportedly higher-level feature, and when the disjunction contains all and only such disjuncts (that is, when the disjunctive property has disjuncts that "exhaustively overlap"), then, Dosanjh maintains, the powers of the disjunction should rather be seen as a proper subset of the powers of whichever disjunct is instanced. And this proper subset relation between powers will, as per usual, be inherited by the tokens of the disjunction and the disjunct types. So, Dosanjh concludes, there is no clear barrier to identifying the type of a realized feature S with the disjunction of its realizer types, since tokens of both types will satisfy the *Proper Subset of Powers Condition*.

My response is to deny that this gambit is available to the reductionist, for two reasons.

First, whether or not the disjuncts of a disjunct are relevantly similar, there is a case to be made that disjunctive features do not have a proper subset of the powers of their disjuncts. Consider the powers of $P \vee Q$ to produce effects when in circumstances C (restrict attention to these powers, for simplicity). For the cases of multiple realization at issue, we can without loss of generality assume that different realizers P and Q cannot be co-instantiated, in which case there will be two ways for $P \vee Q$ to be instanced, and so (at least) two powers which may be exercised in C:

1. If in $C \wedge P \wedge \neg Q$, then E_1; and
2. If in $C \wedge Q \wedge \neg P$, then E_2.

(There may well be others, but that won't matter for making the point.) What powers will P have, in C? It will have at least one of the powers of $P \vee Q$, in C—namely,

If in $C \wedge P \wedge \neg Q$, then E_1.

(Here the conjunct P is redundant, but no matter.) However, P will not have another of the powers of $P \vee Q$, in C—namely,

If in $C \wedge Q \wedge \neg P$, then E_2.

P would have such a power only if it could be both instanced and not instanced in *C* at the same time (powers being relativized to times or temporal intervals), which it can't. So in this case (and more generally), $P \vee Q$ does not have a proper subset of the powers of *P*, or any other of its disjuncts.

Second, the disjunctive strategy is most naturally implemented against the assumption that the reductionist can help themselves to any 'ontologically lightweight' combinations of characteristically physical entities and features, including—all parties agree—any combinations resulting from Boolean operations (with the possible exception of negations). Hence it is that, were it possible to identify a feature *S* with a disjunction of lower-level physical features, that would suffice for *S*'s being ontologically reducible to—that is, identical with—some lower-level physical feature. As Dosanjh (2014) himself notes, "If there are any ontologically innocent combinations of properties, boolean logical combinations are among them. For an important example: when we talk about a disjunction of properties, we are committed to little beyond the properties that serve as the disjuncts" (17). But if the disjunctive property *D* with which *S* is identified is *not* such that its instantiation consists just in the instantiation of its disjuncts, then resources for constructing *D* have gone beyond mere Boolean combination, and it is no longer clear that *D* is appropriately taken to be a lower-level physical property. On the contrary: to the extent that *D*, like *S*, is supposed to satisfy the *Proper Subset of Powers Condition* with respect to *S*'s lower-level realizer, one might rather suppose that *D* (hence *S*) is a higher-level feature, contra reductionism.[13]

3.2.3 Reduction to a metaphysical consequence of lower-level laws?

The third strategy for reductively treating a feature *S* satisfying the conditions in *Weak Emergence* adverts to the metaphysical consequences of the laws governing goings-on at the level of *S*'s base feature—that is, to the fundamental physical laws governing entities and features at the one level that the reductionist thinks exists.[14] Here the reductionist starts by observing that the Weak emergentist is committed to taking higher-level features and laws to be metaphysical consequences

[13] Dosanjh recognizes the concern here, and in response, attempts to undermine the reasons for thinking that higher-level features are irreducible to base features. Most pressingly, he argues that attention to distinctive systems of laws and associated comparatively abstract causal joints need not be seen as supporting higher-level causal autonomy, on grounds that one may "deny that such laws are distinct enough to ground causal autonomy", insofar as "every law statement about a property that satisfies [the *Proper Subset of Powers Condition*] is entailed by law statements about physical properties" (73). This objection is of a piece with, and admits of the same response to, the objection to be next discussed.

[14] Recall that the qualifier 'metaphysical' when applied to consequences of laws is intended to sidestep concerns about the lower-level laws containing vocabulary local to the special sciences: the notions of 'consequence' or 'deducibility' here are metaphysical, not representational.

of, and perhaps even to be theoretically deducible or predictable from, lower-level physical features and laws—after all, it is this which ensures that *S* is physically acceptable. But in that case, the reductionist continues, what prevents *S*, even with its reduced set of powers, from itself being a lower-level physical feature?

Variations on the theme of this concern are common. Consider these remarks by Klee (1984), following Nagel (1961), directed against an account of purportedly physically acceptable emergence on which emergent entities and laws simply involve new relational structures:

> [I]n what sense are these new regularities emergent? To be sure, they may be regularities and structures of a type not found on lower-levels of organization, but it has seemed to some (Nagel 1961, 367–74) that this fact by itself would not justify the label of 'emergent' if they had been predictable on the basis of a thorough understanding of those lower-levels of organization. If the new relational structure which grounds the new regularities could have been predicted on such a basis, then the new regularities could have been predicted and the force of any emergence claim, at least partially, compromised. (46)

Indeed, it might seem practically definitional that (in-principle) theoretical deducibility entails ontological reducibility. As Owens (1989) puts it:

> Reductionism is sometimes expressed as the thesis that the laws of the non-physical sciences can be deduced from those of the physical sciences together with certain bridging generalizations [...]. (63)

To be sure, some higher-level goings-on—in particular, certain complex systems of the sort we will discuss in Ch. 5—might be thought not to be even in-principle deducible from lower-level goings-on, at least given a wide but empirically informed purview of the available resources; but the deeper concern here remains even so. For so long as the nonreductive physicalist maintains that the purported higher-level goings-on are metaphysical consequences of the lower-level goings-on, as they are committed to doing, the concern remains that metaphysically, if not representationally, the higher-level goings-on must be reducible to—that is, identical to—some or other lower-level goings-on, after all. This line of thought seems operative in Kim's (2010) discussion of Shoemaker's subset-of-powers approach to physical realization:

> [Shoemaker's view] strikes me as a very strong form of physicalism, something like type physicalism [...] On an intuitive level, my picture of Shoemaker's scheme is something like this: The fundamental ontological items of this world are physical causal powers [...] These are packaged, or bundled, into properties

> [...] In this picture, where do mental properties find a place? If all properties are bundles of physical causal powers [...] what makes some properties mental and others physical? (110)

Note that Kim is not asking here what makes it the case that some powers are associated with, say, mental properties. Which collections of powers are associated with what features is presumably a matter for scientists (or other relevant practitioners) to decide. Rather, the concern is simply one registering a kind of generic suspicion that if some feature is associated with only physical powers, then that alone suffices to render the feature identical to some physical feature. Hence Kim goes on:

> [I]f physical properties are constituted by purely physical causal powers and mental properties are physically realized in Shoemaker's sense, there seems no good reason not to consider these supposedly mental properties to be physical properties. (111–12)

Morris (2018) endorses this line of thought, and aims to develop it by filling in the operative understanding of physical properties and powers (which Kim leaves open) as satisfying the 'No Fundamental Mentality' (NFM) constraint (as per Wilson 2006; see Ch. 1, §1.4.1):

> Suppose that physical properties and instances of them have all and only powers that are NFM-physical. If all properties that are instantiated are either physical or subset-realized by physical properties and only physical properties, it follows that there are all and only NFM-physical powers. It follows, moreover, that any property or property instance associated with a subset of the powers of a physical property and its instances also has all and only NFM-physical powers. And if properties are the properties that they are in virtue of their powers, it is difficult to see why a property that has all and only NFM-physical powers should not be regarded as NFM-physical. (135)

My response here consists in saying more about how an appeal to laws should enter into the individuation of levels. It is correct that nonreductive physicalists will allow, *qua* physicalists, that all special-science entities and features are metaphysical consequences of lower-level physical goings-on, and that (consonant with *Physical Causal Closure*) the only powers are physical powers (satisfying the NFM constraint, in particular). However, they can reasonably deny that it thereby follows that special-science entities and features are themselves lower-level physical entities or features. In Wilson 2010*b*, I develop this idea by attention to the role that degrees of freedom (DOF) play in individuating scientific entities and features.

To start, note that specification of the law-governed properties and behaviors of lower-level physical entities and features requires all the information needed for the lower-level physical laws to operate, including information about quantum-mechanical DOF and associated values, such as spin and quark color charge. By way of contrast—the nonreductionist can reasonably maintain—specification of the law-governed properties and behaviors of at least some special-science entities and features does not require information about, e.g., quantum-mechanical DOF such as spin or color charge: such DOF are eliminated from these specifications, as unneeded to characterize higher-level goings-on. For example, and as I will develop in more detail in Ch. 5 (§5.2.4) and Ch. 6 (§6.1.1), there are empirical cases to be made that various quantum DOF are eliminated from the DOF needed to specify entities and features of the sort treated by thermodynamics and classical mechanics.

Such eliminations in quantum DOF explain, in part, why special-science entities and features are insensitive to certain micro-level details (and relatedly, make room for such entities and features' being multiply realizable): higher-level macro-entities are typically insensitive to spin-theoretic details, among other quantum features. More importantly for present purposes: that the specification of special-science entities and features does not include quantum-level information means that even though these entities and features are metaphysical consequences of physical laws, it is not appropriate to place them at the physical level: the quantum laws wouldn't know what to do with them! Related considerations, which I expand on in Ch. 5, §5.2.4, provide a basis for seeing how it could be that an entity or feature satisfying the *Proper Subset of Powers Condition*, where each token power is a power of its lower-level physical realizer, could nonetheless not itself be a lower-level physical entity or feature.

As such, one must distinguish two sorts of metaphysical consequences of the lower-level physical laws. First are the entities and features which are consequences in the broadest sense—which, if physicalism is correct, will include any and all special-science (more generally: macro-) entities and features. Second are those metaphysical consequences which retain all the DOF and associated information (pertaining to quantum spin, color charge, etc.) needed for the lower-level physical laws to operate. If there is Weak emergence, then as an empirical matter of fact some entities and features which are metaphysical consequences of the physical laws in the first sense will not be metaphysical consequences in the second sense. Such entities and features will have specifications that fail to include all the DOF and associated information needed for the lower-level physical laws to operate, and so these entities and features will not be appropriately seen as identical to any lower-level physical goings-on—as the nonreductionist maintains.

This line of thought—that laws require certain kinds of information in order to operate, and that among the entities and features that are consequences of

level-L laws, only those preserving the information needed for the level-L-laws to operate are appropriately placed at L—is explicitly encoded in a DOF-based implementation of *Weak Emergence* (see Wilson 2010*b*), and it is some advantage of a DOF-based account that it clearly has the resources to address the concern that theoretical deducibility suffices for or entails ontological reducibility. That said, there does not appear to be any barrier to proponents of other accounts of Weak emergence appealing to the aforementioned general scientific facts about laws and associated DOF by way of responding to the threat of law-consequence reducibility.

3.3 Objection: compatibility with physical unacceptability

Another general line of objection to Weak emergence—again, understood as conforming to *Weak Emergence*—is that a feature *S* satisfying the conditions in the schema might nonetheless be, one way or another, 'over and above' its dependence base feature *P*, in a way rendering *S* physically unacceptable. In what follows I present and address several strategies for establishing this result.

3.3.1 Quiddities

Melnyk (2006, 141–3) supposes that for a feature *S* that cotemporally materially depends on some base feature *P* to be physically acceptable, *S* must satisfy not only the aforementioned 'necessitation' condition, according to which a realized feature is minimally nomologically necessitated by its realizing (i.e., base) feature, but also the 'constitution' and 'truthmaking' conditions, according to which a realized feature is constituted by base features, and truths about realized features are made true by truths about realizing features. Properly restricted to worlds with physical laws of nature similar to those actually governing *P*, Melnyk's suppositions are plausible: it is commonly supposed that *S*'s physical acceptability requires that *S* be necessitated by *P* in worlds containing physical laws;[15] the constitution condition is a clear variant on the theme of complete metaphysical dependence operative in physicalism; and the truthmaking condition might be thought to follow from the

[15] Not everyone discussing necessitation or entailment of features by base features is careful to make explicit that the necessitation or entailment at issue is operative against the backdrop assumption that relevant instantiations of the base features hold fixed the operation of the physical laws. Without this qualification, failures of necessitation or entailment wouldn't clearly count against Weak emergence, at least given that (as some believe), features can enter into different laws and thus have different powers; but whether a feature *P* is associated with a feature *S* in worlds with different laws of nature is neither here nor there for purposes of motivating *S*'s being realized by *P* in worlds with laws relevantly similar to ours.

general physicalist commitment to the lower-level physical goings-on providing a basis, hence a truthmaking basis, for all else.

Does satisfaction of the conditions in *Weak Emergence* guarantee satisfaction of these conditions on physical acceptability? Melnyk is willing to grant that the answer is 'yes', *if* one endorses a 'causal' account of properties along lines of that endorsed by Shoemaker (1980 and 1998), on which features are essentially and exhaustively constituted by the powers they have or bestow. The resulting version of a powers-based account of realization, Melnyk observes, "has the important virtue of meeting the necessitation, constitution, and truthmaking conditions" (144), with "[t]he key move [being] to identify property-instances with something like clusters of causal power-tokens of particular types" (140). Morris (2018) agrees, moreover suggesting that a causal account is required for purposes of guaranteeing the physical acceptability of features satisfying the *Proper Subset of Powers Condition*:

> [H]ow could the subset relation between powers entail that higher-level properties are nothing over and above physical realizers if there is more to higher-level properties than powers [...]? (142)

Granting for the moment the supposition that implementing the schema for Weak emergence requires endorsement of a causal account of properties (features), the problem remains, in Melnyk's view, that for some properties, "it's implausible to identify their instances with clusters of causal power-tokens" (146).[16] The most pressing case among those Melnyk considers has to do with properties whose individuation involves some sort of primitive identity or 'quiddity': the property equivalent of a 'haecceity', or primitive identity, serving to individuate objects or other particulars.[17] By way of illustration, Melnyk refers to Hawthorne's (2001) case of properties which are intuitively distinct but which play the same nomic role (perhaps positive and negative charge are actual such properties). If special-science feature *S* has such a noncausal quiddity, then, it seems, *S*'s satisfaction of the *Proper Subset of Powers Condition* won't guarantee that *S*'s quiddity is constituted by or otherwise 'nothing over and above' *P* (or *P*'s quiddity, as the case may be), or that truths about *S* are made true by truths about *P*, or even that instances of *P*'s type (physically) necessitate instances of *S*'s type. As such, Melnyk suggests,

[16] A similar concern applies to variations on the theme of Shoemaker's account, associated with views on which properties and laws are essentially intertwined (as per Swoyer 1982 and Bird 2001 and 2007), and which would also (at least provisionally) entail that satisfaction of the conditions in *Weak Emergence* would ensure satisfaction of the constitution, truthmaking, and necessitation conditions.

[17] Effectively, a haecceity makes room for the identity and individuation of an entity to float free from any of the entity's features, and a quiddity makes room for the identity and individuation of a feature to float free from any of the feature's powers.

the possibility that features have noncausal quiddities poses a dilemma for the Weak emergentist: either properties may have noncausal quiddities, in which case satisfaction of the *Proper Subset of Powers Condition* doesn't ensure satisfaction of the constitution, truthmaking, and necessitation conditions as required for *S*'s physical acceptability; or quiddities are rejected and a causal account of properties maintained, which ensures *S*'s physical acceptability, but only at the price of endorsing a controversial and (given the seeming possibility of Hawthorne-style cases) not obviously satisfactory account of properties (or features, more generally).

The Weak emergentist can sidestep this dilemma, as follows (see Wilson 2006 and 2011*b*). To start, they can observe that the individuation of scientific features is neutral on the presence or absence of quiddities: in scientific contexts, the occurrence of scientific features, and any truths about such features, does not depend on or otherwise track whether such features have quiddities, much less track how the noncausal quiddities of seemingly distinct features are related. This is true, in particular, for properties such as positive and negative charge, which at some level of abstraction play the same causal role. That physics posits more than one charge property reflects global considerations pertaining to the structure of the laws as requiring that there be two or more distinct properties playing what is in some sense the 'same' role; the posit of primitive quiddities plays no role in this story. As such, the Weak emergentist can reasonably maintain that whether *S* and/or *P* have quiddities, shared or not, is irrelevant to whether *S* satisfies the constitution, truthmaking, or necessitation conditions, and more generally, is irrelevant to whether *S* is physically acceptable; and similarly for artifactual features satisfying the conditions in *Weak Emergence*.

Morris (2018) registers dissatisfaction with this line of thought, saying that "I am not sure that science is exclusively concerned with powers. It is certainly not the case that all sciences characterize all entities and properties in all and only causal terms" (145); here Morris cites Polger (2007), who notes that science may be concerned with, e.g., formal computational properties. My own view is that even formal computational properties are associated with powers. In any case, my suggestion is not that the individuation of scientific properties proceeds only by way of powers; rather, it is that the individuation of such features does not proceed by way of noncausal quiddities.[18] Morris also makes the interesting point that even if scientific features are individuated in a way that is neutral on whether a causal account of properties is correct—or more specifically, for present purposes, on whether scientific features have quiddities, it remains that "physicalism is a comprehensive thesis about the nature and structure of the world [...] the question

[18] In context, Morris also appears to suppose, as is commonly done, that phenomenal aspects of features are not causal; I respond to this sort of concern in §3.3.2.

of whether subset realization suffices for physical acceptability without a causal view of properties immediately reappears as the question of whether the 'scientific' account of the world is the whole world" (145). I take Morris's point, and in response clarify that in my view, quiddities do not enter into the individuation of *any* features, scientific or otherwise, at a world.

Importantly, this response does not require endorsing a causal theory of properties; indeed, this response might be endorsed even by those taking properties to have noncausal quiddities. For the primary reason for endorsing property (more generally, feature) quiddities concerns not a supposed need to distinguish properties with shared roles at a world, but rather a supposed need to identify or individuate properties at or across worlds with different laws of nature, on the assumption that the powers of properties may vary in worlds with different laws (see, e.g., Lewis 1983*a* and Schaffer 2004). Since we do not have any access to noncausal quiddities, if properties are individuated within or at a world even partly by their quiddities, then all bets are off so far as actual property individuation is concerned (see Shoemaker 1980 for a similar epistemological point). As such, those taking properties to have quiddities should agree that the identification and individuation of broadly scientific features within or across worlds with the same (or relevantly similar) laws of nature proceeds by reference to their powers in a way that is neutral on the presence or absence of quiddities.

The upshot is that the possibility of noncausal quiddities poses no threat to the claim that satisfaction of the conditions in *Weak Emergence* is sufficient for physically acceptable emergence.

3.3.2 Phenomenal aspects

A second concern about the physical acceptability of features satisfying the conditions of *Weak Emergence* adverts to the phenomenal or qualitative aspects of certain mental features. As Walter (2010) summarizes the concern, "phenomenal properties just cannot be characterized in terms of their causal role, and thus they cannot be individuated in terms of the causal powers of their bearers to which they contribute" (220). Indeed, it is common to claim that phenomenal aspects of mental features cannot be characterized in terms of causal roles or associated powers, on grounds that such aspects are associated with a phenomenal 'character' or qualitative 'feel' (sometimes described as there being "something that it is like" to have the feature in question) not exhausted by the powers of the associated feature.

Call a feature having phenomenal or qualitative aspects a 'qualitative feature'. As Jonas Christensen notes (pers. comm.), qualitative features needn't be epiphenomenal—an important qualification for present purposes, since on the operative conception of emergent entities and features, these are efficacious

(indeed, distinctively efficacious) as well as distinct.[19] Still, Christensen suggests, attention to qualitative feel provides reason to think that the phenomenality of a feature is not exhausted by the powers of the feature. The intense feel of pain might indeed cause one to cry out and/or behave in certain pain-attending or pain-mitigating ways; nonetheless, the qualitative feel of being in pain is reasonably taken to be an extra feature of reality, in addition to any such powers. Supposing so, then even if a qualitative feature *S* satisfied the *Proper Subset of Powers Condition*, *S* could, in virtue of being associated with a distinctive phenomenal aspect, be over and above its base feature *P*, and hence physically unacceptable.

Moreover, Christensen observes, the means of resisting taking the possibility of quiddities to pose a problem for Weak emergence do not carry over to the case of phenomenal aspects, for two reasons. First, unlike the case of quiddities, scientific theory and practice do appear to be concerned with phenomenal aspects of mental features—for example, pain and how to allay it are part of the subject matter of pharmacology. Second, unlike the case of quiddities, it is commonly thought that if mental features have rich phenomenal aspects that are not fully explained by or (more weakly) somehow metaphysically accommodated in terms ultimately involving lower-level physical powers, then such mental features would be physically unacceptable. So the previous response to the possibility of quiddities doesn't carry over to the case of phenomenal aspects.

I see at least two strategies of response to Christensen's concern. On the first, one maintains that any phenomenal aspects of features there may be are reducible to non-phenomenal representational aspects. Here one might follow reductive representationalists (e.g., Harman 1990, Dretske 1995, Tye 1995, Byrne 2001, Hill 2009; see Chalmers 2004 for discussion) in thinking that phenomenal aspects are reducible to non-phenomenal features of reality, which non-phenomenal features are, in turn, standardly taken to be amenable to treatment in terms of powers.

On the second strategy of response, which I prefer, one rather maintains that phenomenal aspects of mental features are fully incorporated into the powers of these features (compatible with a view on which powers are contingently associated with features, relative to a given set of laws). The suggestion, in other words, is that phenomenal aspects are 'causally loaded'. After all, as Christensen notes, qualitative mental features are plausibly taken to enter into causal relations—in virtue, at least in part, of their phenomenal aspects. In my view, it is reasonable to believe that, as our immediate introspective access to the phenomenal aspects of mental features suggests, any (discernible) differences in phenomenality would result in causal differences—including, if one wants a systematic hook to hang this point on, differences in what sort of qualitative experience the bearer of the

[19] Moreover, as I'll discuss later in the section, good reasons for thinking that phenomenal or qualitative features might be entirely epiphenomenal are in short supply.

mental feature will have and (upon reflecting) take themselves to be having. Call the thesis that phenomenal aspects of a feature are fully incorporated into powers of that feature, perhaps in a way that varies with different laws at different worlds, 'Phenomenal Incorporation'. Given Phenomenal Incorporation, it is reasonable to believe that the powers of a qualitative feature *S* fully incorporate its phenomenal aspects, in such a way that, were each power of *S* to be (on an occasion, etc.) token-identical with a power of its lower-level physical feature *P*, that would indeed suffice for *S*'s being physically acceptable.[20] Similar views have recently been proposed and defended by Langsam (2011) and Mørch (2018) and (2020).[21]

Are there reasons to reject Phenomenal Incorporation? One purported reason might advert to arguments (see Chalmers 1996, 2009, and 2003) for the (suitably ideal) conceivability of zombies, according to which there could be a world physically and functionally (including causally) the same as the actual world, and in particular containing creatures physically and functionally like us, but with a complete absence of phenomenal character. Supposing that zombies are (suitably ideally) conceivable, and supposing also that (suitably ideal) conceivability is a guide to metaphysical possibility, then one might reasonably think that higher-level qualitative mental features are, even if nomologically connected to lower-level physical features and powers, nonetheless wholly distinct from such features and powers in a way at odds with Phenomenal Incorporation. A related scenario (see Block and Fodor 1972 and Shoemaker 1975) involves the (suitably ideal) conceivability of creatures who are physically and functionally identical to us, yet spectrally inverted (with one seeing green where the other sees red, and so on); here again one might see such a possibility, if genuine, as indicating that

[20] Here it is worth recalling that the physicalist is under no obligation to deny that there are phenomenal aspects of natural reality—they merely maintain that any such aspects are metaphysically dependent on suitably complex lower-level physical goings-on. Note also that the supposition that, as per Phenomenal Incorporation, the powers of a Weakly emergent qualitative feature will be identical to powers of its base feature poses no threat to the 'No Fundamental Mentality' (NFM) constraint in the operative conception of the physical (see Ch. 1, §1.4.1), since the dependence base feature will be a feature of a micro-configuration, and as previously disussed, the NFM constraint only rules out mentality as had or bestowed by individual or 'basic' physical entities or features. (Similarly, mutatis mutandis, for variations on the theme of the NFM constraint aimed at making sense of physicalism in infinitely decomposable worlds.) All (non-eliminativist) physicalists will grant that suitably complex physical configurations have features that are identical with or which realize qualitative mental features.

[21] Langsam (2011) claims that conscious states have powers that "flow in an intelligible way from the relevant intrinsic features of consciousness" (4), and Mørch (2020) claims that "phenomenal properties (i.e. properties which characterize what it is like to be in conscious states) have non-Humean causal powers […] in virtue of how they feel, i.e. in virtue of their phenomenal character" (2). Each cites by way of illustration the power of a state of pain to cause the desire to avoid it and the power of phenomenal properties to cause beliefs. I am friendly to these views, though as in the main text, I am inclined to think that one can maintain that the powers of phenomenal mental states incorporate reference to their phenomenal aspects while remaining neutral as regards the metaphysics and modal status of the associated powers.

qualitative mental features float free of lower-level physical features and powers, contra Phenomenal Incorporation.

There is a huge contemporary literature on whether the zombie and/or inverted spectrum scenarios are genuinely (ideally) conceivable, and on whether (suitably ideal) conceivability is a guide to metaphysical possibility.[22] In Ch. 7 (§7.1.2), I will revisit Chalmers's conceivability argument, in particular, in more detail. For now, I aim to lay out two ways of maintaining Phenomenal Incorporation, and the associated viability of Weak emergence, in the face of these sorts of scenarios.

First, and perhaps most importantly, even if the scenarios are taken to be ideally conceivable and metaphysically possible, they do not undercut Phenomenal Incorporation per se; rather, they undercut Phenomenal Incorporation understood as coupled with physicalism. This much is compatible with qualitative features' satisfying Phenomenal Incorporation, but being Strongly emergent (among other options). In that case, however, the scenarios pose no difficulty for the in-principle viability of Weak emergence, for what they establish is not the falsity of Phenomenal Incorporation but rather (at best) the failure of qualitative features to satisfy the *Proper Subset of Powers Condition*. Again: the zombie and inverted spectra scenarios do not show that qualitative aspects float free of powers; they show, at best, that such aspects float free of *physical* powers.

Second, though it is not the present order of business to defend the live possibility of a Weak emergentist account of qualitative features (a possibility to which we will return in Ch. 7), it is worth observing that a Weak emergentist (nonreductive physicalist) has a fairly straightforward reason to deny that such scenarios are genuinely possible. As Perry (2001) correctly notes as regards the zombie case, physicalists (whether of reductive or nonreductive stripe) taking qualitative features to be efficacious will maintain that the absence of such features at a world would entail the absence of the corresponding powers or associated effects at that world, resulting, contra Chalmers's assumption, in a physical or functional causal difference (see Wilson 2002*b* for discussion). Similarly, physicalists accepting Phenomenal Incorporation may reasonably deny that either zombie or inverted spectrum cases are genuinely possible. More generally, independently of whether physicalism is true, it is implausible that inverted spectral differences would fail to make a physical or functional causal difference—one has only to look at a spectrally inverted image of food, for example, to appreciate how comparatively unappetizing the represented contents seem. Given the availability of such physicalist responses, it appears that the zombie and spectral inversion

[22] Re zombies, see the discussion and references in Kirk 2015; re inverted spectrum cases, see the discussion and references in Byrne 2020; for reasons to think that such scenarios import problematic propositional presuppositions, see Hellie 2019*b*; re whether conceivability is the best way to implement 'epistemic two-dimensionalism' (by way of regaining, post-Kripke, some a priori access to modal truths), see the discussion and references in Biggs and Wilson 2020.

scenarios presuppose rather than establish that qualitative aspects float free of physical goings-on (and so are physically unacceptable). Relatedly, only if one antecedently rejects Phenomenal Incorporation, either in general or as compatible with physicalism, will one be inclined to think that there might be worlds physically and functionally identical to ours in which our doppelgangers have mental features entirely lacking in or spectrally inverted with respect to their phenomenal character. As such, the scenarios pose no clear threat to Phenomenal Incorporation, either in general or as compatible with physicalism, and so pose no threat to the claim that satisfaction of the conditions in *Weak Emergence* is sufficient for physical acceptability.[23]

How else might one argue against Phenomenal Incorporation in such a way as to call into question the sufficiency of *Weak Emergence* for physically acceptable emergence? What seems to be required is that it be genuinely possible that qualitative features be epiphenomenal—capable of being instantiated without any powers whatsoever, physical or otherwise. But since, as above, the zombie and inverted spectra scenarios are compatible with the truth of Strong emergence, these scenarios do not establish the genuine possibility of epiphenomenal qualitative features; and given that the having of a qualitative feature crucially involves the having of a qualitative experience—a seemingly causal affair—one might reasonably surmise that no other scenario is going to establish this, either. I conclude that, modulo the presentation of some new and better reasons to reject Phenomenal Incorporation, phenomenal aspects of qualitative features pose no problem for the physical acceptability of features satisfying the conditions in *Weak Emergence*.

3.3.3 Lack of entailment

Several cases have been offered as showing that satisfaction of the *Proper Subset Condition on Powers* by features *S* and *P* is compatible with *P*'s failing to entail or metaphysically necessitate *S*, and so (it is moreover maintained) with *S*'s being physically unacceptable:

- Melnyk (2006) suggests that if *S*'s possession requires having a causal history of some sort (e.g., *being a member of the species Homo sapiens*, or *being a mother*), or requires standing in noncausal (e.g., spatiotemporal) relations (e.g., *being to the right of a rock*), and on the assumption that the base feature *P* is spatiotemporally local, then even if *S* satisfies the conditions in *Weak*

[23] As prefigured, in Ch. 7 (§7.1.2), I will consider and address a more sophisticated line of thought, due to Chalmers, for taking the conceivability of zombies to have anti-physicalist import.

Emergence, *S* might fail to be metaphysically necessitated by *P*, and hence not be guaranteed to be physically acceptable. A similar concern might attach to intentional mental features having 'broad' content: if *S* is a state of believing that water is wet, and the extension of 'water' encodes facts about the chemical composition of the predominantly watery stuff in the environment, and *P* is a given neurophysiological state, then even granting that *S*'s token powers are a proper subset of *P*'s token powers, *P* might not entail *S*.

- McLaughlin's (2007) case pertains to a view of properties (more generally, features) on which they are individuated not just by what effects they may contribute to causing (powers, properly speaking), but also by how they may be caused (as per what Shoemaker calls 'backwards-facing powers'). Insofar as the *Proper Subset of Powers Condition* makes no reference to backwards-facing powers, satisfaction of this condition by *S* with respect to *P* won't ensure that *S* is, as McLaughlin puts it, 'entailed by' (i.e., metaphysically necessitated by) the occurrence of *P*. What is additionally needed to ensure that *P* entails *S*, McLaughlin suggests, is that the backwards-facing powers of *P* entail the backwards-facing powers of *S* (as the *Proper Subset of Powers Condition* does for the forward-facing powers of *S* with respect to *P*); but there doesn't seem to be any clear way of ensuring this. In particular, Shoemaker's (2007) revision of his account of realization to incorporate reference to backwards-facing powers doesn't ensure that *P* even nomologically necessitates *S*, for this revision requires that the token backwards-facing powers of an instance of a higher-level realized feature *S* be a proper *superset* of those of the feature *P* that realizes it on a given occasion. In that case, as McLaughlin observes, the occurrence of *P* will have *fewer* backwards-facing powers than *S*, in which case *P* will not entail *S* (at least, not without further conditions).
- Gibb (2013) offers a case involving circuits in which the powers of a conjunctive feature *S* are a proper subset of the powers of one of its conjuncts *P*, but where (since conjuncts typically do not entail associated conjunctions), satisfaction of the *Proper Subset of Powers Condition* is compatible with *S*'s being physically unacceptable—if, in particular, other of *S*'s conjuncts are physically unacceptable.

A first, general response to these cases proceeds by noting that each case is directed against an account of realization (e.g., that endorsed in Shoemaker 2000/2001) appealing just to the satisfaction of the *Proper Subset of Powers Condition*. The schema for Weak emergence also explicitly involves a cotemporal material dependence condition requiring, among other things, that Weakly emergent features be minimally nomologically supervenient on base features. But in each of the cases, the base feature *P* at issue fails (not just to metaphysically necessitate, but also) to even nomologically necessitate the higher-level feature *S*. (This is

clear for Melnyk's and McLaughlin's cases; the reader will have to take my word for it for Gibb's case, which is a bit too complicated to present in detail here.) Correspondingly, the cases don't in fact satisfy the conditions in the schema for Weak emergence, and hence do not bear on the physical acceptability of features satisfying those conditions.

More specific responses are available for the individual cases:

- As regards Melnyk's cases, a Weak emergentist can maintain that what such cases show is that if the instantiation of a higher-level feature *S* requires that certain spatiotemporal or causal-historical features or facts be in place, then establishing that *S* is Weakly emergent will require identifying a lower-level feature *P* capable of encoding such facts in such a way as to satisfy the dependence condition. For example, features such as *being to the right of a rock* will require realizing features covering the relevant spatiotemporal extent; and if *S* is a historically sensitive species feature—say, *being human*—then if *P* is to minimally nomologically necessitate *S*, then *P* will presumably have to be a spatiotemporally non-local, extrinsic, and presumably massively complex lower-level physical feature.[24] Such spatiotemporal extensions of the base entities and features are familiar from the literature on supervenience as a potential realization relation (to be discussed down the line), where, it is suggested, accommodating extrinsically constituted higher-level features requires taking the physical supervenience base feature to be regional or even global (see, e.g., Horgan 1982, Kim 1984, 1989, and Paull and Sider 1992).[25]
- As regards McLaughlin's case, the Weak emergentist can reject as clearly incorrect a view of properties on which they are typically individuated (in part) by reference to the ways in which they can be caused. For example, scientific properties do not seem to be individuated in this way: it isn't any part of what it is to be an H_2O molecule (even given the actual laws) that such molecules might arise either as the result of natural or artificial processes. If higher-level features are not typically individuated by the ways in which they may be caused, then the fact that higher-level features can typically be caused in more ways than lower-level features poses no threat to either the necessitation of or the physical acceptability of higher-level features satisfying the *Proper Subset of Powers Condition*. Perhaps for such reasons, Shoemaker has since given up individuating properties in terms referencing

[24] Perhaps there are other strategies for encoding such historical facts, but this will do for purposes of illustration.

[25] That non-locality of certain higher-level features might be seen as indicating just that the dependence base feature must be appropriately spatiotemporally extended is also relevant to assessing certain purported cases of Strong emergence (e.g., the case of quantum entanglement); see Ch. 4, note 3 for discussion.

'backwards-facing powers', and reverted to characterizing realization in terms referencing only powers of the usual, forward-facing variety. To be sure, it is plausible that some higher-level properties are individuated in part by reference to how they were in fact caused (as in the case of species kinds); but in that case the Weak emergentist can offer a response similar to that just given to Melnyk, according to which such cases show only that lower-level feature *P* must be spatiotemporally broad enough to ensure satisfaction of the dependence condition.
- As regards Gibb's case, suppose that this case is tweaked to ensure satisfaction of the dependence condition, such that *P qua* conjunct feature (at least) nomologically necessitates *S qua* conjunctive feature. The Weak emergentist might then maintain that given the satisfaction of the *Proper Subset of Powers Condition*, there's no clear reason to think that *S* would be physically unacceptable.

3.3.4 Fundamentally mental powers

Another case aimed at showing that satisfaction of the *Proper Subset of Powers Condition* is compatible with physical unacceptability is due to Baltimore (2013), as follows:

> [A]ccording to Wilson, when the set of causal powers associated with a mental property is a proper subset of the set of causal powers associated with its physical base, the mental property, although distinct from its physical base, will still be physicalistically acceptable [...] There is reason to question, however, the safety of such a retreat. Consider [...] a micro-object at the fundamental level of the micro-macro hierarchy [;] channel the spirit of panpsychism and endow the fundamental micro-object with mentality. However, instead of it having a mental property that is identical with one of its physical properties, suppose that it has a mental property that non-reductively supervenes on one of its physical properties. Suppose further that [...] each causal power associated with the mental property is identical with a causal power associated with its physical base. But the physical base here is a fundamental property and, so, mental causal powers are thereby associated with a fundamental property, which seems physicalistically unacceptable. (20)

This is an interesting case, and I agree with Baltimore that it is not ruled out just by satisfaction of the *Proper Subset of Powers Condition* (or more generally, the conditions in *Weak Emergence*). It is ruled out, however, by the operative physics-based conception of the physical, which as discussed in Ch. 1 (§1.4.1)

incorporates a 'No Fundamental Mentality' (NFM) (or related) constraint, as encoding the historical philosophical motivations for characterizing the physical in such a way as to avoid Hempel's Dilemma. Baltimore considers this response, and replies that the NFM constraint should not be imposed, since it "builds too much metaphysics into the notion of the physical" (15)—ruling out, in particular, that the physical entities might turn out to be as panpsychists suppose. But given the need to preserve the historical contrast between physicalism and panpsychism, and to secure the other advantages of a physics-based NFM account of the physical (see Wilson 2006), I take the import of Baltimore's case to mainly be that of reminding us that nonreductive physicalist implementations of *Weak Emergence* will need to appeal to an account of the physical goings-on suited for physicalist purposes.

3.3.5 Physically unacceptable constraints

Physically acceptable emergence, according to *Weak Emergence*, is ultimately a matter of a higher-level feature *S* having, on a given occasion, only a proper subset of the token powers of the lower-level feature *P* upon which *S* cotemporally materially depends, on that occasion. But presumably not *every* proper subset of powers had by such a *P* corresponds to a distinct higher-level feature. So, one might ask, what distinguishes subsets of powers that are associated with a given higher-level feature from those that aren't? Now, in general it's not the burden of the metaphysician aiming to make sense of emergence to say which higher-level features exist—that's an empirical matter, whether natural or artifactual features are at issue. Still, the Weak emergentist does need to say enough to block answers to the above question having as a consequence that the existence or instantiation of the higher-level feature would be physically unacceptable. As Melnyk (2006) puts it:

> [N]ot just any old subcluster of a given cluster of causal power-tokens constitutes a genuine property-instance [...]. Hence, some further condition must be met by those subclusters that do (see Shoemaker, 2001, pp. 85–6). And it is presumably a task for metaphysics to say what this further condition is [...] the meeting of the further condition must be a purely physical affair [...]. (146)

By way of motivation, Melnyk provides a "frivolous" example, in which a sub-cluster of a given cluster of token powers constitutes a genuine property instance only if it is divinely classified as natural, and to observe that such a scenario would call into question the physical acceptability of the higher-level feature at issue—even were such a feature to satisfy the conditions in *Weak Emergence*.

Before responding, I want to clarify just what the concern here is supposed to be. At least as I read Melnyk, the concern isn't that any property instantiation that might be caused by a divine being would thereby be physically unacceptable. Suppose that a divine being said, "Let there be the big bang and the laws of physics", and in so saying brought our universe into being, to evolve henceforth according to those initial conditions and laws. This sort of scenario would not obviously falsify physicalism—some physicalists are theists—and relatedly, it would not obviously render all natural features (including physical ones!) physically unacceptable. The concern is rather along the following lines: whatever makes it the case that some proper subsets of token powers of a given lower-level physical feature correspond to (instantiated) higher-level features, while other subsets do not do so, had better itself be physically acceptable if the higher-level features are to be physically acceptable; yet satisfaction of the conditions in *Weak Emergence* is silent on why a given higher-level feature *S* has the distinctive power profile it has, and so is compatible (one might think) with the instantiation of a higher-level feature's being, somehow or other, the outcome of a physically unacceptable process.

Melnyk's concern can be more finely pitched by observing that in many cases the existence of higher-level entities and features goes hand in hand with the presence of certain constraints. As the qualifier 'special' suggests, the special sciences concern broadly natural goings-on which are restricted to certain 'special' circumstances—corresponding, e.g., to energies and associated temperatures conducive for the formation of atoms or stable molecules, or where conditions are favorable for life, or where creatures with mentality exist, and so on. Effectively, such restricted circumstances encode the presence of constraints which enter into individuating the associated special-scientific entities, features, and laws as necessarily subject to those constraints. Now, entities and features that can exist only in some circumstances—i.e., only when associated constraints are in place—will be able to do less than, hence will have fewer associated powers than, entities and features that can exist under the restricted circumstances as well as other circumstances; as Lamb (2015) puts it, "constraints are limits on possible states and behaviours" (3). For example, a system of atoms (or associated feature of *being such-and-such system of atoms*) will have more powers than the molecule (or associated feature of *being such-and-such a molecule*) for which it may serve as a dependence base entity (feature), since the system of atoms may exist in, and contribute to the production of effects in, circumstances in which the molecule cannot exist as well as circumstances in which the molecule exists. And one might reasonably suggest that many special-science features which are good candidates for physically acceptable emergence are ones similarly reflecting the holding of certain constraints. That constraints may be operative in the existence and individuation of higher-level entities or features satisfying the *Proper Subset of Powers Condition* provides a way for Melnyk's concern to be pitched—namely as

the concern that these constraints might arise as a result of physically unacceptable processes. Though this concern is a live one, the Weak emergentist has two ways to respond.

First, the Weak emergentist can maintain that, as per the historical dispute between physicalists (materialists) and their Strong emergentist rivals, what distinguishes 'over and above' (i.e., physically unacceptable) features is just that they have powers their base features don't have. So even if some physically unacceptable process were to be operative in making it the case that, e.g., temperatures were within the range needed for stable molecules to exist, this wouldn't in itself show that molecules were physically unacceptable. This response seems to be generally available to proponents of any implementation of Weak emergence.

Second, the Weak emergentist can build into the operative implementation of Weak emergence that any constraints underlying or responsible for the holding of the *Proper Subset of Powers Condition* must occur as a result of physically acceptable processes (or, more generally, as a result of law-governed processes at the level of *P*). In my (2010*b*) DOF-based implementation of *Weak Emergence* (which I will present in more detail in Ch. 5), I include just such a condition.[26] If there is a well-motivated revision of *Weak Emergence* and the associated conception of physically acceptable emergence, I am inclined to see it as involving the inclusion of such a condition on any operative constraints.

3.4 Objection: non-necessity

I turn now to considering certain accounts of physically acceptable emergence which can be seen as challenging the claim that satisfaction of the conditions in *Weak Emergence* is necessary for physically acceptable emergence—notwithstanding that, as argued in Ch. 2, a wide range of accounts of realization clearly do aim to satisfy these conditions. There are three main alternative

[26] See condition 3, below. Since in this implementation I aimed to characterize a Weakly emergent entity (as opposed to just a Weakly emergent feature), the condition more generally requires that any constraints associated with any feature of the entity satisfy the condition:

> *Weak Ontological Emergence* (DOF): An entity *E* is weakly emergent from some entities e_i if

1. *E* is composed by the e_i, as a result of imposing some constraint(s) on the e_i.
2. For some characteristic state *S* of *E*: at least one of the DOF required to characterize a realizing system of *E* (consisting of the e_i standing in the e_i-level relations relevant to composing *E*) as being in *E* is eliminated from the DOF required to characterize *E* as being in *S*.
3. For every characteristic state *S* of *E*: Every reduction, restriction, or elimination in the DOF needed to characterize *E* as being in *S* is associated with e_i-level constraints.
4. The law-governed properties and behavior of *E* are completely determined by the law-governed properties and behavior of the e_i, when the e_i stand in the e_i-level relations relevant to their composing *E*. (Wilson 2010*b*, 292)

approaches on offer, in terms of token identity, constitution, and primitive Grounding, respectively.[27] In what follows, I provide reasons for thinking that each of these alternative approaches to physically acceptable emergence is unsatisfactory.

3.4.1 Token identity

On a token identity approach, a realized feature *S* is token but not type-identical to the base feature *P* upon which it depends, on a given occasion (see, e.g., Davidson 1970, Macdonald and Macdonald 1995, Ehring 2003, and Robb 1997). Such an approach entails that every token power of *S*, on an occasion, is identical to a token power of base feature *P*, on that occasion, and hence avoids token-level causal overdetermination between *S* and *P*. It doesn't gain *S*'s (token-level) ontological autonomy (distinctness), but one might think that this isn't as important as gaining *S*'s reality and efficacy. Moreover, the token-identity theorist maintains, the causal autonomy characteristic of emergence can be accommodated by taking *S* to be distinctively efficacious in virtue of falling under a distinctive type.

One concern with this approach is that appeal to non-identical types as the ground of the causal autonomy of *S* with respect to *P* reintroduces a threat of higher-level causal overdetermination, and associated threat of exclusion. As Ehring (2003) puts the concern: "Since mental types are not identical to physical types (because of multiple realizability) even if mental tokens are identical to physical tokens, there are no causes of physical effects that are efficacious in virtue of mental property types" (364). To gain *S*'s causal autonomy, the proponent of a token identity account must provide an account of the relation between *S* and *P*'s associated types, and show that the associated means of gaining autonomy does not reintroduce problematic overdetermination.

Ehring (2003) aims to provide an account of the relation between types on which problematic overdetermination is avoided. To start, he takes *S* and *P* to be tropes—particularized properties, such as *this redness*, or *that complex configuration of charges*—and their associated types to be collections of resembling tropes. He then argues that *S* and *P*'s types are related as part to whole. Here the order of the part/whole relation is reversed from Shoemaker's (2000/2001) and Clapp's (2001) understanding: for Shoemaker and Clapp, a realized type is part of each

[27] A fourth line of objection to the necessity claim appeals to supervenience, with the general idea being that it suffices for physically acceptable emergence that a higher-level feature asymmetrically metaphysically supervenes on lower-level physical features. However, since appeals to supervenience in this context are primarily aimed at distinguishing physically unacceptable emergence from any sort of physically acceptable relation, I postpone discussion of supervenience as a basis for metaphysical emergence until the next chapter.

of its realizing types; for Ehring, a realized type is a whole, having as parts the subclasses (of resembling tropes) of its realizing types.

What is interesting for present purposes is that Ehring takes appeal to a type-level 'proper subset of powers' strategy to be required by the token identity theorist to establish the requisite causal autonomy without inducing problematic overdetermination. He first motivates the view for the determinable type *red* and its associated determinate shade types:

> It seems clear that the class of red tropes as a whole, the type "red", has certain causal powers. [...] [W]e are still left with the question of how the causal powers of this class as a whole are related to the causal powers of the subclasses of each determinate shade of red. I believe the answer is that the causal powers of the type "red" are those exactly similar causal powers shared by each of these subclasses. [...] For any causal power of a shade of red not matched by an exactly similar causal power belonging to each of the other shades of red, "red" lacks any such power. (374)

Here the powers of the type *red* are a proper subset of those of each of its constituent determinate types. Ehring takes similar considerations to indicate that the powers of mental types are a proper subset of the powers of their realizing physical types. Macdonald and Macdonald (1995) also plausibly implement a subset-of-powers-based strategy at the level of types, for they take mental state types to be relevantly analogous to determinables, and as previously, a case can be made that determinable types have a proper subset of the powers of their realizing determinate types.

Such hybrid approaches, combining token identity of features with a proper subset relation between powers of associated types, are problematic, however, for a reason that I have already observed. On the hybrid view, *S*'s type does not have powers that differ between its realizer types; but a token of *S*'s type can, when identical with a token of *P*'s type, have such powers. Hence token feature *S* can have powers that *S*'s type doesn't have. But it arguably makes no sense for a token feature to have more powers than its type, at least if types are supposed to track similarities among associated tokens. If a token feature has more powers than a given type, then that is itself compelling reason to think that the token feature is not of that type, or so it seems to me.

Avoiding this difficulty requires that the proper subset relation between powers hold at the level of tokens as well as types—that is, that the *Proper Subset of Powers Condition* be imposed. More precisely, it requires imposing the *Proper Subset of Powers Condition* if the account is to be a version of nonreductive physicalism (physically acceptable emergence). Alternatively, the proponent of a token identity account could endorse reductionism at both token and type levels. At the end of the

day, token identity accounts of realization either do not establish the ontological and causal autonomy of higher-level features, and so are not really accounts of physically acceptable emergence; or else impose the *Proper Subset of Powers Condition*, and so are not really token identity accounts.

3.4.2 Many-one constitutive mechanism

In Ch. 2, I discussed constitutive mechanism accounts of realization of a higher-level feature of a system by a lower-level micro-structural feature, where the latter feature encodes information about the system's parts and provides a causal mechanism for the implementation of the causal role and associated powers associated with the higher-level feature. Such accounts, I argued, plausibly ensure satisfaction of the *Proper Subset of Powers Condition,* and more generally are properly seen as implementing the schema for Weak emergence. There I also noted, however, that a second variety of constitutive mechanism approach to realization, notably endorsed by Gillett in his 2002*b*, 2002*a*, 2016, and elsewhere, rather takes realization to be a many-one relation between the many token features of individual parts of a system and a token feature of the system as a whole.

One might think that Gillett's version of mechanistic realization would not satisfy the *Proper Subset of Powers Condition*, on pain of saying, implausibly, that every token power of a mechanistically realized feature is identical to a token power of some or other individual part of the system having the feature. And indeed, Gillett does charge those endorsing what he calls the 'Subset View' with this implausible consequence. In fact, however (and echoing remarks made in Ch. 1, note 11), there is a case to be made that a many-one mechanistic approach to realization is committed to satisfaction of the *Proper Subset of Powers Condition*. As Craver (2001) emphasizes, a given constitutive mechanistic explanation does not proceed simply by citing the features of individuals which compose or otherwise serve as a cotemporal material base for a given macro-entity. Indeed, Craver goes further and says that "In fact, the φ properties of mechanisms are not really mere properties [of individuals] at all; they are the activities of and among the entities in the mechanism" (59). More specifically, he observes:

> An analytic account for a mechanism is not just a list of entities and activities; it is a description of a mechanism. And that description involves, in addition to a list of entities and activities, a description of how they are organized together actively, spatially, and temporally in S's φ-ing. Specifying the mechanistic role of some component X, accordingly, involves describing how X is organized with the other entities in S such that it contributes to S's φ-ing. (61)

In metaphysical terms, Craver's point tells against an understanding of mechanistic realization according to which a single higher-level feature of a system is somehow realized by—not a list, but a plurality—of co-instantiated features of individual parts of the system. Nor would Gillett deny that in order for mechanistic realization to occur, the parts and associated features need to be appropriately spatiotemporally and causally integrated. But if a many-one account of mechanistic realization is feasible only if the many parts and their many features are appropriately integrated, there must needs be some metaphysical locus of this integration; and the natural thing to say here is that this locus is something like a micro-structural feature of the sort at issue in the previously discussed mechanistic accounts, which again plausibly conforms to the conditions in *Weak Emergence*.

That said, Gillett expresses three concerns with an account of realization or Weak emergence involving satisfaction of the *Proper Subset of Powers Condition*: first, that such an account is committed to realizing and realized features being instantiated in the same entity (as per what he calls a 'flat' rather than a 'dimensioned' approach); second, that such an account is unable to accommodate qualitative differences between features of macro-entities and their composing micro-entities (as when, e.g., a diamond, but not a carbon atom, is hard); third (and related to the previous two concerns), that such an account fails to accommodate the idea that individual entities and features 'realize' configurations and associated micro-structural features.

I address each concern, in turn. First, the *Proper Subset of Powers Condition* is neutral on whether realized and realizing features are had by the same entities. To be sure, some (including myself, in my 1999) take the condition to apply in cases of what I there called 'same-subject necessitation'. But as I now maintain (see Wilson 2011*b* and 2015*b*), this assumption is dispensable. Second, as for qualitative differences between features (and associated powers) of, e.g., carbon atoms and diamonds: these are understood and accommodated, in the first instance, as reflecting differences between features of individuals (e.g., carbon atoms) and features of micro-configurations (e.g., complex arrays of carbon atoms). Against the backdrop assumption of physicalism, qualitative differences are accommodated not via an inter-level realization relation, but rather via the various intra-level relations that are operative in bringing individuals at a given level together into configurations at that level. Third, one may deny that the latter sorts of intra-level many-one spatiotemporal and causal relations holding between, e.g., (many) lower-level physical individuals and their features and (one) lower-level micro-configuration and its features should be subsumed under the rubric of 'realization'—most importantly, because this sort of intra-level relation does not provide a basis for making sense of higher-level reality.

The upshot is that accounts of realization ensuring the *Proper Subset of Powers Condition* do not have the problematic implications Gillett attributes to them—a result he should embrace, since he is arguably also committed to satisfaction of this condition, if constitutive mechanisms are to be properly integrated.

3.4.3 Constitution

On Pereboom's (2002) account of "robust" nonreductive physicalism—i.e., of physically acceptable emergence—a higher-level feature *S* is neither type- nor token-identical with the lower-level physical feature *P* upon which it depends; and contra both the *Token Identity of Powers Condition* and the *Proper Subset of Powers Condition*, *S*'s token powers are "irreducible to" powers of *P*: "robust nonreductive physicalism affirms various token-diversity claims for mental causal powers" (500). Such a view will clearly make sense of *S*'s ontological and causal autonomy. But how are *S*'s token-irreducible powers supposed to avoid problematic causal overdetermination (as discussed in Ch. 2, §2.1.1) while retaining compatibility with physicalism? According to Pereboom, this is because *S*'s powers are "constituted" by *P*'s powers, in a way piggybacking on the notion of token feature constitution:

> *Token Power Constitution*: The causal powers of a token of kind *F* are constituted of the causal powers of a token of kind *G* just in case the token of kind *F* has the causal powers it does in virtue of its being constituted of a token of kind *G*. (504)[28]

The notion of constitution of one token feature by another is broadly primitive, but is (as per Pereboom 2011) to be grasped as relevantly analogous to the 'made up of' relation holding between one particular and another (e.g., a statue and a lump of clay). The account of feature constitution, coupled with *Token Power Constitution*, is intended to motivate taking the powers of a realized feature *S* to be, while irreducible to, still nothing over above the powers of *P*. As Pereboom puts it:

> [Though *S*'s token powers are irreducible to *P*'s] there would be a sense in which the token causal powers of *S* would be "nothing over and above" the token causal powers of *P* […] *S*'s causal powers would nevertheless be "absorbed" or "swallowed up" by *P*'s causal powers. But there are importantly distinct modes of this sort of absorption: identity and constitution without identity. […] token mental causal powers are wholly constituted by token microphysical causal powers. (503–4)

[28] See also Pereboom and Kornblith 1991, 131.

(I have changed Pereboom's notation for continuity with my discussion.)

But such appeals to token feature and power constitution do not establish that rejection of the *Token Identity of Powers Condition*, hence *Proper Subset of Powers Condition*, is compatible with physicalism. 'Constitution' is a term of art, applied mainly (as Pereboom notes) to objects. Where token features are at issue, and where conformity to physicalism is presumed, 'constitution' is usually just another name for 'realization'. But as previously argued, standard accounts of realization presuppose satisfaction of the *Token Identity of Powers Condition*. The expression 'in virtue of' entering into the account of token power constitution is also a term of art, compatible with many underlying relations, including identity (satisfying the *Token Identity of Powers Condition*) and the determinable/determinate relation (satisfying the *Proper Subset of Powers Condition*).[29]

Pereboom offers further considerations in support of irreducible mental powers' being compatible with physicalism and with the avoidance of overdetermination, but these also fail to establish his case. In re compatibility with physicalism, he says that "correlated with the possibility of this sort of constitutional explanation is the fact that the existence and nature of token higher-level causal powers would be predictable in principle from their microphysical constituents together with the laws governing them" (504). But if the powers of *S*, at either the type or token level, are not identical with the powers of *P*, what guarantees that the powers of *S* would be so predictable? Perhaps such predictability could be guaranteed if every power of *S* was type-identical (though token-distinct) with a power of *P*, but this understanding appears to give rise to pervasive causal overdetermination.[30]

Pereboom resists this conclusion, saying that "no competition arises in the case of mere constitution":

> For if the token of a higher-level causal power is currently wholly constituted by a complex of microphysical causal powers, there are two sets of causal powers at work which are constituted from precisely the same stuff [...] and in this sense we might say that they coincide constitutionally. (505)

To the extent that I understand why constitutional coincidence blocks overdetermination, however, this is because (the relevant sort of) coincidence would entail identity of the token powers had by features of the 'stuff'. Pereboom acknowledges "that they now coincide in this way might tempt one to suppose that these causal powers are token-identical, but [...] there is a good argument that they are not" (505).

[29] One might be tempted to interpret this 'in virtue of' talk in terms of a primitive notion or relation of 'Grounding' or 'Ground'. This would be a mistake, for reasons I'll discuss in the next section.

[30] See Audi 2012*b*, 14 for discussion of a similar concern.

Here Pereboom is referring to a line of thought presented in his (2002). There he first suggests that a token higher-level mental state *S* might be multiply realizable, and that if so *S* would not be identical with its token base state *P*:

> Suppose that *S* is realized by a complex neural state [*P*]. It is possible for *S* to be realized differently only in that a few neural pathways are used that are token distinct from those actually engaged. [...] [I]t is evident that this alternative neural realization is itself realized by a microphysical state *P′* that is token distinct from *P*. It is therefore possible for *S* to be realized by a microphysical state not identical with *P*, and thus *S* is not identical with *P*. (503)

He then continues:

> [T]his reflection would also undermine a token-identity claim for mental causal powers—should they exist—and their underlying microphysical causal powers. For if the token microphysical realization of *S* had been different, its token microphysical causal powers would also have been different. We therefore have good reason to suppose that any token mental causal powers of *S* would not be identical with the token microphysical causal powers of its realization. (503)

As I discuss in Wilson 2015*b*, the Weak emergentist can respond to this line of thought in either of two ways. First, they can deny that tokens of higher-level features are ever multiply realizable. What is (fairly) uncontroversially true is that *types* of higher-level features are multiply realizable, in the sense that different tokens of the realized type can be realized by tokens of different lower-level types. Pereboom's case, above, for taking token higher-level feature *S* to be multiply realizable isn't compelling, and may be rejected. Compare: could that very instance of red, currently realized by an instance of scarlet, have been realized by an instance of burgundy? One might reasonably judge not, and continue to reasonably judge not even supposing the alternative shade to be only "slightly" different from the original. One may similarly reasonably deny that *S* (as opposed to another token of *S*'s type) could be realized by a base feature other than *P*, whether this alternative feature is very different from *P*, or only different with respect to "a few neural pathways" or other lower-level physical details.

Second, the Weak emergentist can grant that a token feature *S* may be multiply realizable, but maintain that in such a case, *S*'s token powers are relativized to realizers (or occasions). Pereboom presupposes that *S* has its token powers essentially; but why think this? As he later observes, "stable tokens [...] often retain their identity over certain changes in their constitutions and configurations" (529). One might similarly maintain that token feature *S* can retain its identity across changes in its realizers and associated token causal powers. The *Proper Subset*

of Powers Condition requires only that *S*'s token powers on a given occasion be numerically identical with the powers of the lower-level physical feature realizing it on that occasion; hence the *Proper Subset of Powers Condition* can be satisfied even if *S* is token multiply realizable.

3.4.4 Grounding

It has recently been suggested that complete metaphysical dependence can or should be understood in terms of a primitive relation or notion—what I've called ('big-G') 'Grounding', to distinguish it from historically common schematic uses of the terms 'grounding' (or 'ground')—which is supposed to be operative in any context where some goings-on or facts hold 'in virtue of' some others (see Fine 2001, Schaffer 2009, Rosen 2010, Audi 2012*a*, Raven 2015; see Bliss and Trogdon 2016 for an overview and other references). Though proponents disagree about which formal features are characteristic of Grounding, it is commonly supposed that Grounding is (at a minimum) asymmetric and irreflexive, in which case Grounded goings-on are ontologically autonomous (distinct) from Grounding goings-on.

One of the main suggested applications of Grounding (see, e.g., Schaffer 2009, Rosen 2010, and Dasgupta 2014) is as providing a basis for formulating non-reductive physicalism, as follows:

> *Physicalism* (Grounding): All broadly scientific goings-on are Grounded in lower-level physical goings-on.

The suggestion is supported by appeal to the following form of argument:[31]

1. Physicalism is the thesis, schematically speaking, that all broadly scientific goings-on are completely metaphysically dependent on ('nothing over and above', etc.) lower-level physical goings-on.
2. The operative notion of complete metaphysical dependence cannot be successfully characterized in semantic/representational, epistemic, or purely modal (i.e., supervenience-based) terms.
3. No other non-primitive approach to characterizing complete metaphysical dependence is available.

∴ The operative notion of complete metaphysical dependence in physicalism should be characterized in terms of Grounding.

[31] See Schaffer (2009, 364), Rosen (2010, 111–12), and Dasgupta (2014, 557).

Coupled with the usual supposition that Grounding is asymmetric and irreflexive, such an argument might be seen as offering an account of the realization relation making no reference to the *Proper Subset of Powers Condition*, and so as suggesting that satisfaction of the conditions in *Weak Emergence* is not necessary for emergence of a physically acceptable variety.

As I've argued in a series of papers (Wilson 2014*a*, 2016*d*, 2018/2016), there are many ways in which the lines of thought leading to the posit of Grounding and its application to physicalism go wrong. A full discussion of my concerns here would take us too far afield; in what follows I simply sketch two concerns which are relevant to the question at hand, and direct those interested in further details to my other work on the topic.

First, as I note in Wilson 2014*a*, arguments like that above for positing Grounding are unsound, since premise (3) is false. As will already be clear from discussion in this and previous chapters, during the past several decades philosophers working on physicalism have identified and explored many specific relations of the sort that might enter into formulating physicalism, either in general or as applied to a specific case, which relations are explicitly assumed to go beyond representational, epistemic, or merely modal notions. These small-'g' grounding relations (as I call them in my 2014*a*)—including type and token identity, functional realization, constitutive mechanism, the part/whole relation, the determinable/determinate relation, the proper-subset-of-powers relation, and so on—serve, against the backdrop assumption that the physical goings-on are the sole fundamenta, to characterize diverse forms of physical dependence (some reductive, some nonreductive) in an explanatory and illuminating way. Given the availability of these properly metaphysical alternatives, there is no route from the failure of epistemic, representational, and/or merely modal (i.e., supervenience-based) conceptions of metaphysical dependence to a primitive Grounding-based understanding of this notion (much less to one tracking just metaphysical dependence of a nonreductive variety). Other motivations for positing primitive Grounding have been advanced, but as I have argued, these are also uncompelling.[32]

[32] One stated motivation appeals to unity considerations (see Schaffer 2016*b* and 2016*a*, Berker 2017), with the basic idea being that Grounding is needed in order to formally and otherwise unify the small-'g' relations. But as I and others have argued (see Wilson 2014*a*, 2016*d*; Koslicki 2012, 2016), the small-'g' relations are not formally or otherwise unified. Even if they were unified, more would be required to establish that such unity motivates a generic worldly posit, as opposed to merely motivating a schematic or generic concept. To highlight just one case in point, determinable features formally and otherwise unify determinate features, but it is nonetheless commonly supposed either that determinables don't exist or that they are reducible to a disjunctive or other combination of determinates (see Wilson 2017 for discussion). My own view is that determinables do exist and may be as fundamental as their determinates (see Wilson 2012), but my point here is that this requires argument: unity considerations alone aren't enough to motivate a metaphysical posit, much less a primitive one.

Second, even granting that Grounding exists, the lack of non-stipulative content associated with this primitive posit renders it incapable in itself of shedding any illuminating light on the notion of metaphysical dependence it is introduced as explicating.[33] Most importantly, appeal to Grounding in itself provides no basis for a physicalist response to the problem of higher-level causation, according to which the effects of a higher-level entity or feature are causally overdetermined, since already brought about by base-level entities or features. As we've seen, considerable nonreductive physicalist action concerning this problem has involved identifying one or other specific relation (e.g., functional realization, constitutive mechanism, the part/whole relation, the determinable/determinate relation, and so on) whose holding would illuminate how, exactly, some higher-level entities or features might be both physically acceptable and distinctively efficacious in a way avoiding problematic (double-rock-throw-type) overdetermination. By way of contrast, the holding of a primitive Grounding relation says nothing—at least, nothing non-stipulative—about whether and how causal overdetermination might be avoided as between higher- and lower-level entities and features, or even about whether a Grounded entity has any powers at all; hence in itself this relation fails to provide a basis for responding to the main barrier to making sense of physically acceptable emergence. Correspondingly, a Grounding-based approach to Weak emergence does not satisfy the criterion of illuminating accommodation—that is, does not provide a clear and illuminating basis for accommodating the appearances of metaphysical emergence in natural (straightforward, default) fashion.

All this said, nothing prevents a proponent of Grounding from stipulating or otherwise imposing a further condition as doing that work. One option would be to stipulate that in general or in the case at hand a Grounded feature has powers which are moreover a proper subset of those of its Grounding feature, on any given

Another stated motivation appeals to a purported need for Grounding to fix the direction of priority of instances of the specific relations, since (as is illustrated by the dispute between Monists and atomists over whether the Cosmos *qua* whole is prior to or posterior to its proper parts) instances of these relations do not come with a built-in direction of priority. But as discussed in Wilson 2014*a*, 2016*d*, 2018/2016, and in progress*b*, what more is needed is not a primitive pointer (i.e., Grounding) but rather a specification of what is or serves as fundamental. For example, in the case above: given that the Cosmos is the sole fundamental entity, proper parts of the Cosmos are nonfundamental, whereas given that the atomic parts are the sole fundamenta, mereological fusions of the parts (including the Cosmos) are nonfundamental. For another example: given that the basic physical goings-on are the sole fundamenta, then lower-level configurations of the basic physical goings-on are nonfundamental, as are higher-level goings-on standing in various specific relations (functional realization, constitutive mechanism, the part/whole relation, the determinable/determinate relation, and so on) to such lower-level configurations of physical goings-on. Priority between nonfundamenta, where it exists, is then a function of various considerations, including considerations registering the nature of the nonfundamenta as determined by how they small-'g' stand to the fundamenta, and considerations registering how the nonfundamenta are related to each other.

[33] See Bryant 2020 for a recent broadly pessimistic assessment along these lines of Grounding's usefulness for formulating physicalism.

occasion. Another option would be not so different from the one that Horgan (1989) and Lepore and Loewer (1987 and 1989) implement—namely, to require that, in addition to the holding of the (schematically characterized) realization relation at issue, certain difference-making conditions be in place, as reflected in the truth of certain subjunctive conditionals. Either way, however, it would turn out that Grounding *qua* realization would conform to the schema for Weak emergence—in which case that relation, even granting that it exists, poses no threat to taking satisfaction of the conditions in *Weak Emergence* to be necessary for physically acceptable emergence.

3.5 Concluding remarks

In Ch. 2, I considered and responded to what is perhaps the most pressing concern with the viability of physically acceptable emergence—namely, that associated with Kim's problem of higher-level causation—with the key components of the response being the identification of the *Proper Subset of Powers Condition* lying at the heart of the schema for Weak emergence, and the related appreciation of a second way (different from that associated with the having of a new power) in which a higher-level feature might be distinctively efficacious—namely, by having a distinctive power profile, tracking difference-making considerations or a comparatively abstract system of laws or causal joint.

In this chapter, I have considered a wide range of objections to the viability of Weak emergence, according to which satisfaction of the conditions in the schema is compatible with anti-realism, reductionism, or physical unacceptability, or is not necessary for physically acceptable emergence; and I have argued that each objection admits of one or more responses available to proponents of any of the diverse implementations of the schema.[34] In two cases, additional responses are available which rely on features specific to either a determinable-based or a DOF-based implementation of *Weak Emergence*. In particular, a determinable-based account (and perhaps also a DOF-based account) provides a non-stipulative basis for ruling out the Weak emergence of conjuncts from lower-level conjunctions, and a DOF-based account explicitly includes a condition specifying that the holding of any constraints entering into or responsible for the holding of the *Proper Subset of Powers Condition* must be a matter only of physically acceptable (more

[34] Additional objections and responses to the viability of Weak emergence were more conveniently treated in Ch. 2. See my §2.3.1 response to Kim's (2006) claim that realized features inherit all, not just some, of the token powers, my note 27 response to Morris's (2013) attempted motivation of Kim's claim (as per what Morris calls 'Full Inheritance'), and my note 35 response to Morris's (2011) concern that the motivations for taking satisfaction of the *Proper Subset of Powers Condition* to block problematic overdetermination hinge on a problematic analogy to cases of part/whole overdetermination.

generally, base-level) processes. Again, my sense is that if any of the considered objections requires tweaking the conditions in *Weak Emergence*, it is that (due to Melnyk) pertaining to the need to require the physical acceptability of any operative constraints associated with the higher-level feature and its distinctive power profile. The addition of such a condition would be straightforward, and hardly fatal to the overall approach; but again, this is a choice point, and how exactly a Weak emergentist chooses to respond to this or other objections may depend on further of their commitments. In any case, these results collectively indicate that *Weak Emergence* is not just a viable and indeed attractively robust means of accommodating physically acceptable emergence, but that this schema moreover captures what is core and crucial to (and when suitably filled in, what is necessary and sufficient for) such emergence.

4

The viability of Strong emergence

I now turn to a project similar to that of the last chapter, only as directed at the schema for Strong emergence. Recall that in Ch. 2, I provided prima facie reasons for thinking that satisfaction of the conditions in the schema for Strong emergence is core and crucial to a feature's being metaphysically emergent from—cotemporally materially dependent on, yet ontologically and causally autonomous with respect to—a lower-level feature, in a way incompatible with physicalism, given that the base feature is physically acceptable. Again, the schema is as follows:

> *Strong Emergence*: What it is for token feature *S* to be Strongly metaphysically emergent from token feature *P* on a given occasion is for it to be the case, on that occasion, (i) that *S* cotemporally materially depends on *P*, and (ii) that *S* has at least one token power not identical with any token power of *P*.[1]

Henceforth, by 'Strong emergence', I will intend to refer to the view (or the phenomenon, understood in light of the view) that conformity to the conditions in *Strong Emergence* is core and crucial—and when sensibly filled-in, both necessary and sufficient—for emergence of the sort incompatible with physicalism.

In this chapter, I consider and respond to a range of objections that have been or could be made to the viability of Strong emergence, so understood. These objections fall into four main categories, according to which satisfaction of the conditions in *Strong Emergence* is incompatible with scientific theory or practice (§4.1); is impossible, since any purportedly novel powers of Strongly emergent features are inherited by (or "collapse" into) base features (§4.2); is compatible with physical acceptability (§4.3); or is not necessary for emergence of a physically unacceptable variety (§4.4). As I'll argue, each of the objections admits of one or

[1] Recall that the novelty of powers at issue in *Strong Emergence* is fundamental novelty, but that the qualification that the novelty is fundamental rather than nonfundamental (stemming merely from various less complex goings-on entering into certain aggregative relations) is not needed, given that *P*, as a feature of a lower-level configuration (plurality or structural aggregate), has any such nonfundamental (merely aggregative) powers. Recall also that for a feature's being Strongly emergent it plausibly suffices that the conditions are satisfied on at least one occasion by at least one instance of the feature in worlds with laws relevantly similar to the actual laws; for continuity with the schema for Weak emergence I stick with the schema expressed in terms of occasions.

Metaphysical Emergence. Jessica M. Wilson, Oxford University Press (2021). © Jessica M. Wilson.
DOI: 10.1093/oso/9780198823742.003.0004

more responses that could be endorsed by any proponent of Strong emergence, however the schema is implemented. And as in the case of Weak emergence, upon occasion an additional response to an objection is available which relies on a specific implementation of the schema. In particular, certain attractive responses appeal to a fundamental interaction-relative account of Strong emergence, of the sort I have previously endorsed.

4.1 Objection: incompatibility with scientific theory or practice

Strong emergentists maintain that some higher-level features have new powers, associated with higher-level laws, forces, or interactions that are as metaphysically and scientifically fundamental as the fundamental physical powers, laws, forces and interactions. In terms of forces, for example: a Strongly emergent feature might be supposed to have a new power to produce a fundamental 'configurational' force—that is, a force present only upon the occurrence of certain lower-level configurations (pluralities or structural aggregates). When such forces are present, the operative forces are a combination of physical and Strongly emergent forces. The powers at issue might include powers to contribute to causing some special-scientific effect, but in any case will, for reasons brought out in Ch. 2 (§2.2.1), include powers affecting the motion of lower-level physical entities, conforming to a common understanding of Strong emergence as involving the nomological possibility of 'downward causation'. Several concerns with Strong emergence, so understood, have to do with its being in tension with the content or practice of scientific theories; I address these in turn.

One such concern is that Strong emergence is incompatible with contemporary physics. In response, McLaughlin (1992) convincingly argues that Strongly emergent configurational forces or interactions (of the sort posited by the British Emergentists, in particular) are compatible with the laws and conservation principles of physics (see also Horgan 1993 and Papineau 2001). For example, Newton's second law of motion, $F = ma$, is neutral as regards which component forces enter into the net force F. As Kane (1993) notes:

> $F = ma$ is used to compute the motion of an object, given *any* force F on the object. And specific classical forces have been discovered, such as gravity with $F = \frac{G_N mM}{r^2}$ [...] . Hamilton's or Lagrange's equations are equivalent to $F = ma$ in a different formulation. In quantum theory there is an analogous structure. The Schrödinger equation [...] is like $F = ma$. It holds for any Hamiltonian. Specific forces lead to specific Hamiltonians. (2–3)

Accordingly, this law is compatible with the operative forces' including a fundamental configurational force. Similarly, McLaughlin (1992) observes, for the

contemporary descendant of Newton's law, Schrödinger's equation $H\psi = ih\frac{\partial\psi}{\partial t}$, into which is inserted the Hamiltonian H specifying the energies of the state (forces and energies being inter–translatable[2]):

> It is not that British Emergentism is logically incompatible with nonrelativistic quantum mechanics. It is not. Schrödinger's equation could be the fundamental equation governing motion in a world with energies that are specific to types of structures. (54)

Nor are fundamental configurational forces (more generally: interactions) or energies incompatible with conservation laws, such as the relativistic principle of conservation of mass-energy. As McLaughlin notes:

> [C]onfigurational forces need not involve any violation of this principle. [. . .] Configurational forces could involve various compensating shifts in mass and energy that maintained conformance to the principle of mass-energy. (74)

A seemingly different defense of the scientific compatibility of Strong emergence is offered by Cucu and Pitts (2019), who suggest that were any form of interactionist dualism to be true (as regards minds, in particular), this would violate energy conservation—but that such a violation would not be problematic, but rather just what one would empirically expect. As they put it:

> The view defended here is that the best response to the energy conservation objection—the response that reflects an understanding of the relevant theoretical physics—is what has been called the 'conditionality response' (Pitts 2020) that energy is conserved when and where minds do not act on bodies, but is not conserved when and where minds act on bodies. (100)

Cucu and Pitts's understanding of the relevant conservation law is one, however, which builds in that the conserved energy is physical (is 'physically conserved'); hence their view is not really in tension with the previous response, according to which mental or other Strongly emergent energies or forces would not violate conservation laws, neutrally understood. In any case, their interesting survey of scientists—including Descartes, Newton, and Euler—who appear to have countenanced the possibility of nonphysical energies or forces provides further support for thinking that Strong emergence is not incompatible with physical science.

A second concern is that even if there is nothing in-principle problematic about adding fundamental configurational forces/interactions or energies to the mix of

[2] See Wilson 2007 for discussion and some derivations.

physical forces/interactions or energies entering into the operative equations of motion, such posits are incompatible with scientific practice, or relatedly, with a 'naturalist' outlook, according to which metaphysical investigations and accounts should be consonant with such practice. As I discuss in Wilson 2002*a*, however, scientific theorizing itself provides a blueprint for how conservation laws might enter into the warranted posit of Strongly emergent configurational interactions and associated powers, laws, and features. In the 1930's, the law of conservation of mass-energy appeared to be violated in nuclear β-decay interactions. Rather than accept the apparent violation as genuine, physicists posited a new fundamental interaction—the weak nuclear interaction—as carrying away the missing energy (see Greiner 1996 for discussion). Nuclei are composite entities; hence evidently scientists have no problem with positing fundamental configurational forces/interactions. As it happens, the nuclear interactions are now understood as ultimately due to interactions between sub-nuclear entities, but the point remains: the posit of fundamental configurational forces/interactions and associated powers, which come into play only at certain levels of comparatively complex organization, is compatible with scientific practice, and hence with a naturalistic approach to metaphysical theorizing.

A third concern is that, even granting that there is nothing in-principle problematic about there being fundamental configurational novelty (of powers, forces/interactions, laws), at present science provides no empirical support for such posits. This is the considered judgement of McLaughlin (1992), who maintains that there is "not a scintilla of evidence" in favor of there being Strongly emergent features; Ladyman and Ross (2007) register a similar opinion. Strictly speaking, such a lack of evidence would not constitute an objection to the viability of Strong emergence per se, as opposed to its actual applicability. Even so, it's worth noting that the claim that we presently don't have any reason to think that there is any Strong emergence is overstated. That certain historical candidates for Strong emergence (e.g., the sorts of chemical interactions that Broad discussed) now admit of lower-level physical explanations doesn't show that all such candidates have been so explained. Indeed, as we will see in later chapters, there continues to be considerable debate over whether phenomena such as consciousness and free will are Strongly emergent, in ways in keeping with a naturalistic methodology.[3]

[3] Two other recently discussed cases are worth mentioning here. First, Silberstein and McGeever (1999, 187–9) offer quantum entanglement as a case of emergence (see also Humphreys, 1997, 216); second, Hendry (2010) argues for "the ontological emergence of molecular structure with respect to quantum mechanical systems of nuclei and electrons interacting via Coulomb forces" (220; see also Hendry 2017). As discussed in Ch. 2 (§2.2), the operative accounts of emergence in these discussions are reasonably supposed to be of the Strong variety. That said, I postpone detailed discussion of these purported cases of such emergence for a future occasion (see the to-do list at the end of Ch. 9), in part because quantum entanglement is both poorly understood and not among the target cases of inter-level emergence, and in part because certain difficulties attach to the stated motivations for each purported case of Strong emergence. The considerations Hendry offers in in his (2010) appear to presuppose that

A final concern about the compatibility of Strong emergence and scientific theory and practice has to do with the usual understanding of such emergence as involving commitment to the nomological possibility of downward causation. Wouldn't the efficacy of Strongly emergent properties vis-à-vis physical or physically acceptable effects violate *Physical Causal Closure*, according to which every lower-level physical effect has a sufficient purely lower-level physical cause? And isn't *Physical Causal Closure* widely accepted? Yes, and yes. However, *Physical Causal Closure* is not a principle of contemporary physics (though no doubt many physicists, and scientists more generally, believe it). The acceptance of this principle is rather a constraint on *physicalist* theorizing (which motivates, in particular, reductive and nonreductive physicalist approaches to the problem of higher-level causation); hence that Strong emergentists deny it is not in itself a strike against their view.

4.2 Objection: collapse

The second line of objection aims to show that the conditions in *Strong Emergence* are never jointly satisfied, as per what Taylor (2015) evocatively calls the 'collapse' objection, versions of which have been raised and/or addressed by several philosophers (see, e.g, van Cleve 1990, Kim 1999, O'Connor 1994, Wilson 2002*a*, Francescotti 2007, Howell 2009, Taylor 2015, and Carruth 2018). The general concern is that Strong emergence makes no sense, since any purportedly Strongly emergent features or associated powers 'collapse', one way or another, into the lower-level base features upon which they depend, undermining the supposed ontological and causal autonomy of the emergent features.[4]

it suffices for molecular structure to be Strongly emergent that the structure (or associated features) is not determined solely by the locally interacting parts. In particular, he says "the explanation of why molecules exhibit the lower symmetries they do would appear to be holistic, explaining the molecule's broken symmetry on the basis of its being a subsystem of a supersystem (molecule plus environment). This supersystem has the power to break the symmetry of the states of its subsystems without acquiring that power from its subsystems in any obvious way. That looks like downwards causation" (215–16). However, as I observed in Ch. 3 (§3.3.3) when considering the bearing of spatiotemporally non-local higher-level features on satisfaction of the conditions in *Weak Emergence*, such non-locality might be seen as indicating just that the dependence base feature must be appropriately spatiotemporally extended. (See also note 8 of this chapter for another observation in re Hendry's view.) The main consideration offered by Silberstein and McGeever in their (1999) is that in the case of entanglement "the parts exhibit a holistic (emergent) correlation property possessed by the system but not locally carried by the separate parts" (187), but as discussed in Ch. 1 (§1.4.2), the emergence at issue here concerns the possibility of features that are novel with respect to features of underlying configurations, not just with respect to features of individuals entering into the configurations. Again, there is more to say here but given the present target this will have to do.

[4] A related concern is that such collapse, combined with the supposed physical unacceptability of Strongly emergent features and powers, threatens the physical acceptability of the base features (see Howell 2009). Since Howell offers this objection in support of a supervenience-based approach to physically unacceptable emergence, I postpone discussion of his concern until §4.4.2.

As discussed in Baysan and Wilson 2017, there are two main versions of the objection, to be discussed in the following sections. After presenting these versions and noting, along the way, certain difficulties with previous responses, I offer more promising responses, drawing on Wilson 2002*a* and Baysan and Wilson 2017.

4.2.1 Collapse via power possession

The first route to the collapse objection is one according to which an intuitive way of assigning powers to features entails that any purportedly new power of a Strongly emergent feature *S* will be inherited by its base feature *P*. The concern here underlies what Kim (2006) calls the "critical" question of emergence:

> *M*, as an emergent, must itself have an emergence base property, say *P*. Now we face a critical question: if an emergent, *M*, emerges from basal condition *P*, why cannot *P* displace *M* as a cause of any putative effect of *M*? [...] If causation is understood as nomological (law-based) sufficiency, *P*, as *M*'s emergence base, is nomologically sufficient for it, and *M*, as *P**'s cause, is nomologically sufficient for *P**. It follows that *P* is nomologically sufficient for *P** and hence qualifies as its cause. (558)

Given that, as we are assuming, a feature's powers are a matter of what effects the having of that feature can contribute to causing, when in certain circumstances, the threat to the viability of Strong emergence is clear. For the Strong emergentist standardly supposes that the cotemporal material dependence of a Strongly emergent feature *S* involves, at a minimum, the base feature's being nomologically sufficient for *S*; moreover, nomological sufficiency (in the circumstances, as per usual) is transitive. Consider, then, any power of *S* to contribute to causing an effect *E* in circumstances *K*. If causation is a matter of nomological sufficiency in appropriate circumstances, if *P* is nomologically sufficient for *S* in *K*, and if *S* is nomologically sufficient for *E* in *K*, then *P* will also be nomologically sufficient for *E* in *K*, and so also have the power to contribute to causing *E* in *K*, ruling *S*'s Strong emergence out of court. The upshot is that any supposedly novel powers of *S* will 'collapse' into those of *P*.

Insofar as the line of thought here depends on certain assumptions about causation, one might wonder whether the Strong emergentist can respond by rejecting a view on which nomological sufficiency in the circumstances is sufficient for causation, or by denying that causation is transitive, or by denying that causation can be (as with *P* and *S*) cotemporal. Such responses are unsatisfactory, however. To start (following Hall 2004), accommodation of many intuitive cases of causation requires a notion of causation as 'production', involving nomological sufficiency in

the circumstances; moreover, other accounts of causation seem likely to introduce similar or other difficulties.[5]

Most importantly, even if it is possible to block taking *P* to cause *S* in cases of Strong emergence, there would remain a case to be made that *P* inherits any powers of a feature *S* that at least nomologically cotemporally depends on *P*. Here O'Connor's (1994) presentation of the following "strong objection" to a powers-based account, which he credits to Carl Ginet, is apropos:

> If an emergent property is a necessary consequence of certain base-level properties (as is implied by the supervenience [i.e., dependence] condition), then its instantiation is one of the potentialities of that set of properties. But then are not the further potentialities of this emergent property also a subset of the total set of potentialities of the base properties, in virtue of the necessary connection between the base properties and it? These further potentialities are simply potentialities of the base properties at one remove. And now one is led to wonder why we might ever think to postulate an emergent property at all, since it provides no explanatory gain over an account which excises the mediating link by taking the "further" potentialities as directly tied to the base properties. This objection implies, in effect, that the features of supervenience and novel causal influence are incompatible. (98)

The deeper collapse objection raised here, as well as in Kim's (1998) framing of the objection, does not hinge on the supposition that Strong emergence can be seen as a causal relation, but rather just on the supposition that *P* cotemporally necessitates *S*, with at least nomological necessity. Such necessitation alone suggests that anything that *S* can do in circumstances *K* is also something that *P* can do in circumstances *K*, in which case there is no way for *S* to have a novel power, and so no way for it to be Strongly emergent.

O'Connor (1994) offers a response to the deeper collapse concern, but it is less than satisfactory. He suggests that if *P* is taken to inherit *S*'s powers, then the

[5] Most saliently, if causation is counterfactual dependence (the other main category of causation that Hall identifies), such that a power is associated with a feature only if the associated effect is counterfactually dependent on the feature, two difficulties ensue. First, if the counterfactual dependence concerns the token instances of *S*, *P*, and *E*, then it might be reasonably thought that if *S*'s power to cause *E* reflects *S*'s being necessary in the token circumstances for *E*, then *P* also has the power to cause *E*, in being necessary in the token circumstances for *S*. So collapse remains. (Note also that the previous line of thought is not affected by taking counterfactual dependence to be relative to a given contrast class.) If the necessity rather attaches to the types at issue, then a different problem arises—namely, that any higher-level feature with multiple dependence bases will be deemed Strongly emergent, including multiply realized features that are intuitively physically acceptable. It shouldn't be that easy to falsify physicalism! And while it would be less costly to deny that causation must be transitive or to require that it be diachronic, these denials are both overly committing and ad hoc.

lower-level physical laws would have a "very odd complexity, involving tacked-on disjuncts to cover the special cases" (98). Effectively, O'Connor's suggestion is that collapse would entail that lower-level entities (e.g., atoms) would interact with each other in a uniform way until entering into a complex aggregate, at which point they would (collectively) start doing "quirky" things, and that such discontinuous behaviors are better explained by positing Strongly emergent features. O'Connor's response presupposes that complex behavior, if it is to be physically acceptable, must be smoothly aggregative. That's incorrect, however, for physicalists (reductive or nonreductive) are happy to allow that quirky behavior can come about simply as a result of complexity (as with, e.g., chaotic nonlinear systems).

4.2.2 Collapse via lower-level dispositions

A second version of the collapse objection focuses on the question of when a feature is appropriately placed at the presumed lower level of physical goings-on. An early version of this objection is registered by van Cleve (1990), who after arguing that physically unacceptable emergence represents the best option for making sense of dependent but irreducible higher-level mental features, says of Broad's 'in-principle failure of deducibility' account:

> There is one more point about Broad's account that needs to be discussed. It could be objected to what has so far been said that there is simply no room for the concept of an emergent property, since for any property *P* of any whole *w*, there will always be properties of the parts from which *P* may be deduced. For example, is it not true of sodium that it comes with chlorine to form a whole having such-and-such properties, including its odor and anything else one might have claimed to be emergent? And from such properties of the parts, may not all properties of the whole be deduced? The answer, of course, is yes; but it is also clear that if properties of this sort are admitted in the "supervenience base," the doctrine of anti-emergence [...] becomes completely trivial. (223–4)

Taylor (2015) develops this line of thought, observing that Broad took sodium chloride to be Strongly emergent (that is, to have fundamentally novel powers, etc.), on grounds that from complete knowledge of the properties of sodium and chlorine in isolation, or in compounds different from that associated with sodium chloride, one could not deduce that salt will dissolve in water. But, Taylor argues, it seems that dispositional properties are among the features that can be had by the components "in isolation", in which case the characteristic features and associated powers of sodium chloride will be deducible, after all:

> This case of emergence 'collapses' when [. . .] dispositional properties are included among the micro-level properties. [. . .] For example, one of the characteristic properties of sodium chloride is its solubility in water. Accordingly, sodium has the following dispositional property: to generate a compound that is soluble in water when combined with chlorine into sodium chloride. In Broad's terms, this property is a property of sodium 'in isolation'. [. . .] The emergent features of the whole $R(A, B, C)$ can obviously be deduced from complete knowledge of the features of the parts A, B, and C and the knowledge that they are arranged as a whole $R(A, B, C)$, so long as the features of the parts include these dispositional properties. (736)

Taylor sees a general problem here for accounts of Strong emergence:

> [C]ases of emergence presuppose a distinction between micro-level and macro-level properties. For any purported case of emergence, there are properties that prima facie belong to the micro level, but if they are included in the micro level then the purported emergent fails to meet a necessary condition for emergent autonomy. I call these problematic properties collapse-inducing properties because when they are included in the micro level, the purported emergent effectively 'collapses', and yet it seems arbitrary to exclude them. [. . .] This is the problem of *collapsing emergence* (or, for short, *the collapse problem*).
>
> (732–3, emphases in the original)

Again, the problem such a 'dispositional move' poses for the viability of Strong emergence is clear. Both van Cleve and Taylor focus on Broad's 'failure of deducibility criterion', but as above, the intended import of this criterion is to track the fundamental novelty of a Strongly emergent feature, as reflected in such a feature's having powers not had by the lower-level physical features upon which it depends (or, for that matter, by any other lower-level physical feature). And notwithstanding that the dispositional features of the 'isolated' lower-level entities at issue in van Cleve's and Taylor's discussions are, to use O'Connor's (1994) terminology, at various 'removes' from either P or S (understood, as per usual, as features of lower-level configurations or of macro-entities, respectively), nonetheless such dispositions call into question the intended fundamental novelty of a Strongly emergent feature. Here again O'Connor's (1994) discussion is useful in highlighting the deeper concern at issue, which he credits to Sydney Shoemaker (pers. comm.), according to which one can always insist that purportedly Strongly emergent features are in fact "further (hitherto undetected) micro-properties" which are manifested only in certain complex circumstances.

O'Connor (1994) and van Cleve (1990) offer responses to this version of the collapse objection, but these responses are again less than satisfactory.

O'Connor maintains that it would be "implausible" and ad hoc to posit micro-properties that make their presence known only in highly complex systems: "the only motivation one could have for postulating [such a] micro-property is a very strong methodological principle to the effect that one is to avoid emergentist hypotheses at all costs" (98). But it seems clear that in general, dispositions "make their presence known" only when certain conditions are in place, and sometimes such conditions might well involve highly complex states of affairs (here again cases of complex nonlinear but presumably physically acceptable behavior are relevant); so the mere fact that micro-dispositions would manifest in complex circumstances is not enough to show that the collapse-inducing suggestion is either implausible or ad hoc.

Van Cleve suggests that restricting the base features to those that are manifested in non-emergence-engendering combinations might do the trick:

> Clearly, some sort of anti-triviality stipulation is required. Perhaps the required work can be done by Broad's phrase "taken separately and in other combinations," for one could plausibly refuse to regard the property "forming a whole with such-and-such features when combined with chlorine" as a property of sodium taken separately. (223)

But it is not clear that Broad's qualification provides a basis for plausibly refusing to regard the property "forming a whole with such-and-such features when combined with chlorine" as a property of sodium "taken separately"; for it is commonly assumed (see, e.g., Martin 1996) that dispositions can be had by individuals even when the dispositions aren't being manifested: a vase can be fragile, for example, even if it is never broken. Taylor considers another response—namely, to require that lower-level features be non-dispositional. As she correctly notes, however, this would be overly restrictive, since many uncontroversially lower-level physical features—e.g., having a mass of 5 g—are to some extent dispositional.

4.2.3 Three responses to the collapse objection(s)

I now present three more satisfactory responses to the collapse objection(s); again, see Wilson 2002*a* and Baysan and Wilson 2017 for further discussion.

Direct vs. indirect powers

Perhaps the simplest line of response is one distinguishing between direct and indirect having of powers. Here the Strong emergentist grants that while in cases of Strong emergence there is a loose sense in which *P* or other lower-level physical features inherit *S*'s purportedly new power (due either to *P*'s being nomologically

sufficient for *S*, or to *P* or some other lower-level features' being disposed to give rise to *S*), in a stricter sense *S*'s novel power is not had or manifested by lower-level features in the same direct or immediate way as they are had or manifested by *S*. Notwithstanding that *P* (at least nomologically) necessitates *S*, *P* has the power at issue only in that *P* is a nomologically sufficient precondition, in the circumstances, for *S*, which is the more direct locus of the power. Similarly for lower-level dispositions of isolates which are even further removed from *S* than is *P*: granting that there are such dispositions and that these in some sense refer to *S* and its novel power, the notion of disposition here is again simply that of a precondition or precursor of a feature (namely, *S*) that more directly has the power in question.[6]

Such a strategy seems intuitively well-motivated, given the Strong emergentist understanding of lower-level physical goings-on as being precisely such nomologically sufficient preconditions for Strongly emergent features. There are, moreover, two ways to substantiate the intuition and associated strategy.

First, the Strong emergentist can appeal to an analogy to temporally extended causal chains: even if each link in the chain is, in the circumstances, nomologically sufficient for the next link, one can nonetheless distinguish more and less direct causes of the end result; and the mere fact that, say, a person lights (or could light) a fuse leading to the explosion of some fireworks doesn't entail that the explosion isn't a novel phenomenon, or that there is any but an indirect sense in which that person has the power to produce such an explosion. Similarly, the Strong emergentist can maintain that in cases of Strong emergence the base feature *P* is metaphysically, if not temporally, antecedent to *S* in the chain of feature instantiations potentially leading to the effects associated with *S*'s novel power.

Second, the Strong emergentist can appeal to an analogy to sets and subsets to make the notion of the cotemporal yet indirect having of a power more precise, as reflecting different circumstances associated with *S* and with *P* with respect to the having of the power at issue. As is uncontroversial, powers are individuated, in part, by the circumstances in which they manifest and contribute to the production of a given effect; but just as we can distinguish between a set and its subsets at a time, there seems to be no in-principle reason why we cannot distinguish between different sets of circumstances associated with a single temporal interval (instantaneous or extended). In particular, the Strong emergentist can say that *P* has the power to contribute—nomologically, if not causally—to the production of *S*, in circumstances *K* which do not include the presence of S. In virtue of having

[6] Note the contrast here with Shoemaker's suggestion as discussed in Ch. 3 (§3.1) that in cases of physical realization (Weak emergence), certain of the powers of a realized feature may be had in a more 'direct' way than as had by the realizing feature. As previously discussed, given physicalist acceptance of the theses of *Physical Causal Closure* and, relatedly, the *Token Identity of Powers Condition*, there isn't any room for such a distinction between ways of having powers to get a grip. The Strong emergentist rejects both theses, however, and so can accommodate this distinction.

this power, *P* indirectly has the power to contribute to causing anything that *S* can cause. By way of contrast, *S* has at least one power to contribute to the production of a given effect *E*—namely, its novel power—directly, which power is manifest in circumstances *K′* which, whatever else they might be or contain, do not include the absence of *S*.

Perhaps the main concern with the direct/indirect having strategy is that there may be some indeterminacy in what counts as direct (as opposed to indirect) having or manifestation of a power, just as there might be indeterminacy as regards which link in a causal chain is most temporally proximal to a given effect. Here the Strong emergentist has two responses. First, they can maintain that, as per usual, the presence of indeterminacy or borderline cases needn't undermine the usefulness of a given distinction. Second and relatedly, they can avail themselves of one or other of two strategies for accommodating indeterminacy in properly metaphysical (as opposed to merely semantic or epistemic) terms: first, a metaphysical supervaluationist approach along lines of Akiba (2004), Barnes (2010), Barnes and Williams (2011), and others; second, a determinable-based approach along lines of Wilson (2013*a* and 2016*a*), and recently applied by Bokulich (2014), Wolff (2015), and (Calosi and Wilson, 2019 and forthcoming) to the case of quantum metaphysical indeterminacy. In Ch. 6 (§6.1.3) I'll discuss these accounts in more detail.

Powers relativized to fundamental interactions

A second response to the collapse problem appeals to an independent way of sorting powers, based in the notion of a fundamental interaction, which makes room for higher-level features to have powers that are in some sense new, as Strong emergence requires (see Wilson 2002*a*). It is a scientific truism that powers are metaphysically dependent on one or more fundamental interactions. The power of being able to bond with an electron, in circumstances where one is in the vicinity of a free electron, is grounded in the electromagnetic (or electroweak) interaction, as opposed to the strong nuclear or gravitational interactions. The power of being able to fall when dropped, in circumstances where one is poised above Earth's surface, is grounded in the gravitational force (or its geometric correlate), as opposed to the other fundamental interactions in operation. The power of being able to bond with other atomic nuclei in a stable configuration is grounded in the strong nuclear interaction, as opposed to the electromagnetic, weak, or gravitational interactions. The power of being able to sit on a chair without falling through it is grounded (at least) in the gravitational and the electromagnetic interactions. And so on. In providing a metaphysical basis for the powers bestowed by properties, fundamental interactions systematically explain vast ranges of natural phenomena. As Auyang (1999) puts it, in discussing the currently accepted fundamental interactions:

> There are four fundamental interactions. Gravity holds our feet on earth and the earth in orbit; it is responsible for the large-scale properties of the universe [...] Electromagnetism binds electrons and nuclei into atoms and atoms into molecules; it is responsible for all physical and chemical properties of solids, liquids, and gasses. The strong interaction binds quarks into nucleons and nucleons into atomic nuclei. The weak interaction is responsible for the decay of certain nuclei. (46)

Similarly, Greiner (1996) says:

> All the known interactions that occur in nature can be reduced to four interactions between material particles. Listed in order of decreasing strength, these are: the strong (nuclear) interaction, electromagnetism, the weak (nuclear) interaction, and gravity. (1)

The metaphysics of fundamental interactions, treating the nature of such interactions and how they serve as a basis for powers, is an underdeveloped area of research, and a full exploration of these interesting issues would take us too far afield. Here I will limit myself to saying just enough about these issues to motivate the interaction-based strategy of response to the collapse objection.

To start, the notion of an interaction is a contemporary generalization of the notion of a force: whereas forces are pushes or pulls (or component contributions thereof), now commonly seen as ultimately involving particle exchanges, interactions may involve not just forces but other sorts of interactions, such as particle creations and annihilations. As in the case of (what used to be called) fundamental forces (e.g., gravity, electromagnetism), talk of a fundamental interaction is shorthand for talk of token interactions of a given type, that do or may occur in certain circumstances, and which are at least partly constituted by the presence of features (e.g., charge) lawfully associated with the interaction. As above, certain interactions are deemed fundamental, in the sense of providing a metaphysical basis for all other interactions and associated phenomena. Which criteria are operative in deeming a given form of interaction fundamental is again a large question; for present purposes what is most crucial is that there is an operational test for the introduction of a novel fundamental interaction—namely, that the posit of the interaction is needed to 'balance the books' as regards various quantities (e.g., energy) which are taken to be conserved. Hence it was, as discussed previously, that the weak nuclear interaction was introduced in response to seeming violations of conservation laws associated with radioactive decay.

Though the operative test for positing a new fundamental interaction does not hinge, it seems, on any particular metaphysical account of such interactions or how these provide a basis for powers, it may nonetheless be worth registering certain options on this score. To start, it is common to take a given fundamental

interaction to either be or be associated with a specific collection of fields. If fields are understood as objects (or some other kind of entity), then it might be natural to see them as having features and associated powers of their own, in which case one part of the answer to the question 'how are powers grounded in fundamental interactions?' would be that certain powers—ones plausibly deemed fundamental—are grounded (schematically speaking) in fundamental interactions in virtue of being associated with features of fundamental fields. The question would remain of how exactly the powers of ordinary entities (objects, systems, and other particulars) are grounded (again, schematically speaking) in fundamental interactions; and here one answer might be that these nonfundamental powers are second-order powers of fields: powers of fields to contribute to producing powers of ordinary entities or their ultimate non-field substantial components (e.g., protons and electrons). A somewhat more lightweight metaphysical variation on this theme would interpret fields not as objects (entities), but as collections of comparatively fundamental features and associated powers at spacetime points or regions; here again one might take powers of ordinary entities or their constituents to be second-order powers of powers of spacetime points or regions. And of course, as per usual, there here remain the usual options for understanding talk of powers in more or less heavyweight terms. Independent of further metaphysical details, however, given that the claims that distinct fundamental interactions exist and serve as a foundational basis for the spectrum of diverse powers of ordinary objects are claims in unassailably good scientific standing, a Strong emergentist is within their rights to speak of a feature's having (or not having) a power, relative to a given set of fundamental interactions.

Of course, physicalists of whatever stripe think that fundamental physical interactions are the only fundamental interactions there are, while (as McLaughlin 1992 emphasizes) the Strong emergentist thinks that, in addition, there are one or more nonphysical 'configurational' fundamental interactions. Strong emergentists can thus grant that, taking both physical and nonphysical fundamental interactions into account, a base feature *P* has every power an associated Strongly emergent feature *S* has; but also coherently maintain that such an *S* will have powers that are 'new' relative to those powers of *P grounded only in fundamental physical interactions*. Such a conception clarifies the sense of (fundamental) novelty at issue in the *New Power Condition* in *Strong Emergence*, making explicit that this novelty—hence Strong emergence itself—is interaction-relative, along the following schematic lines:

> *Interaction-relative Strong Emergence*: Token feature *S* is Strongly emergent from token feature *P* relative to the fundamental interactions in a set $\{F\}$ just in case (i) *S* cotemporally materially depends on *P*, and (ii) *S* has at least one token power that is not identical with any token power of *P* that is grounded only in fundamental interactions in $\{F\}$.

Condition (i) again minimally specifies cotemporal material dependence, understood as involving both (physical) substance monism and the minimal nomological supervenience of higher-level on lower-level features, while condition (ii) refines the *New Power Condition* in *Strong Emergence*, making explicit that the sense of 'new' at issue adverts in part to a fundamental interaction that is new relative to some specified set. Again, the use of 'grounded in' at issue here and elsewhere is intended as schematic for some or other specific metaphysical relation, to be further determined (see Wilson 2014*a*). For purposes of characterizing Strong emergence in a way that appropriately contrasts with physicalism, one specifies that the interactions in $\{F\}$ are the fundamental physical interactions.

An interaction-relative implementation of *Strong Emergence* is clearly in the spirit of the original British Emergentist suggestion that Strong emergence involves what "we may call 'configurational forces': fundamental forces that can be exerted only by certain types of configurations of particles" (McLaughlin 1992, 52). And it makes room for there to be Strong emergence in the face of the collapse objection(s). To start: even if, taking all fundamental interactions into account, features of the composing system in some sense inherit all the powers of any features they cotemporally nomologically necessitate, it remains that higher-level features may be associated with powers that are new, in not being grounded *only* in the set of physical fundamental interactions. Properly relativized, the novel powers of Strongly emergent features do not collapse. Relatedly, relativizing powers to fundamental interactions provides a principled basis for distinguishing dispositions expressing mere preconditions for the occurrence of Strongly emergent features from those that are more directly implicated in the having of the novel powers at issue.

One might be concerned that, as with conceptions of emergent features as "surprising", or with Taylor's alternative (2015) conception in terms of what is (perhaps only currently and contingently) scientifically unexplained, a relativized conception of physically unacceptable emergence will fail to track anything metaphysically interesting or 'joint-carving'. *Interaction-relative Strong Emergence* doesn't have this problem, however: new fundamental interactions are interesting and joint-carving, if any natural phenomena are. Another concern might be that the conception requires realism about fundamental forces or interactions. It is unclear, however, just what is supposed to be problematic about such notions.[7] As previously, fundamental interactions are plausibly understood as second-order dispositions of fundamental fields (namely, dispositions to give rise to further dispositions of non-field entities), and dispositions are not just familiar but

[7] See Wilson 2007 for a defense of the reality and irreducibility of Newtonian forces; as least some aspects of this defense would carry over to defense of the reality and irreducibility of fundamental forces and/or interactions.

moreover admit of more or less metaphysically lightweight interpretations. In any case, participants in the debates over physicalism or Strong emergence typically are happy to take a (fallibilist) realist stance towards the posits of science, including any fundamental interactions there might be; so even bracketing or leaving open further metaphysical details, the Strong emergentist is within their rights to appeal to such interactions in framing their account.

Strongly emergent objects

The third available response to the collapse problem is motivated by the thesis that features have their powers derivatively on the powers of their bearers, as suggested by Baysan 2016 (see also Baysan and Wilson 2017). Drawing on this idea, the Strong emergentist might maintain that the novelty of powers at issue in Strong emergence can be understood as always involving the coming-into-existence of a new object (more generally, entity; I stick with 'object' for continuity with Baysan's discussion), suited to be the bearer of *S* and its novel power, and which object is different from the object(s) bearing base feature *P*. Call this 'the new object strategy'. Indeed, though it is common to assume that what powers an entity has are a matter of what features (and associated powers) it has, one might rather maintain that the association of powers with features is derivative on the association of powers with objects. As Baysan (2016) puts it:

> What do we mean when we attribute powers to properties? [...] Being knife-shaped has the power to cut bread—conditionally on being instantiated with certain other properties, of course. When we attribute this power to the property of being knife-shaped, do we really mean that the property itself has this power? Unless we want to identify properties with bundles of powers, I don't think that we have any good reason to give an affirmative answer to this question. Properties don't cut bread. Their bearers might. To generalize, properties don't (literally or fundamentally) have powers; their bearers do. (386)

Accordingly, the Strong emergentist can maintain that the novel powers associated with a Strongly emergent feature are in the first instance powers of a novel object. Such a view provides the basis for a principled response to the collapse objection(s): in cases of the Strong emergence of a feature *S*, *S*'s novel power presupposes the coming-into-existence of a new object, different from the bearer(s) of *P*. Since, on this approach, powers are derivative on the objects having the powers (and associated features), *P* would inherit *S*'s power only if *P* were born by the same object as *S*; but again, the Strong emergentist can reasonably maintain that this is incompatible with *S*'s having the novel power at issue, on grounds that new powers require new objects. Effectively, the new object strategy turns the collapse objection on its head: given that Strong emergence requires a

novel power, and given that powers of features are derivative on powers of their bearers, Strong emergence requires a novel object to have the power, which is then associated with emergent feature *S*. Indeed, the Strong emergentist can implement the strategy even if they don't agree with Baysan that the powers of features are always derivative on the powers of objects. They can simply maintain that new objects are required to be the bearers of any *fundamentally novel* powers or features there might be.

One might wonder whether the implied commitment, on the new object strategy, to the existence of distinct but spatiotemporally coincident objects is problematic. Here I think that the Strong emergentist is within their rights to shrug their shoulders. To start, the prima facie appearances of emergence encourage such a pluralist commitment: special-science entities appear to be spatiotemporally coincident with lower-level physical configurations, individual persons appear to be spatiotemporally coincident with their bodies, and so on. To be sure, there remains controversy over how best to treat seemingly coincident objects, a debate which is often characterized as over the nature of material constitution (see Wasserman 2010 for an overview); but antecedent to the identification of some specific difficulty with Strongly emergent coincident objects, the Strong emergentist who endorses the new object strategy can maintain that their view constitutes one among the available pluralist options for accommodating material constitution.

Another concern with the new object strategy is that it might be seen as avoiding the collapse objection only by giving rise to an 'explosion' objection—namely, by committing the Strong emergentist to a form of substance dualism, contra the traditional supposition that emergence of whatever variety is supposed to conform to substance (more specifically: physical) monism.

Here the Strong emergentist has two main lines of response. First, they may grant that a Strongly emergent object counts as a new (type of) substance, and moreover one that is in some sense nonphysical, but maintain that notwithstanding the traditional characterization of emergence as a form of substance monism, what is most important is that viable forms of such emergence suitably contrast with views on which the additional substances or associated subjects of Strongly emergent features are immaterial or otherwise very different from physical substances. A commitment to Strongly emergent objects or associated 'substances' doesn't entail anything of this sort. Indeed, the assumed cotemporal material dependence of Strongly emergent entities and features on lower-level physical entities and features is typically offered as a basis for contrasting this view with serious (e.g., Cartesian) forms of substance dualism.

Second, the Strong emergentist can deny that from the mere positing of a new object (entity) a new substance is thereby posited. To start, the claim that all and only objects are substances is controversial, and may be rejected. For example, Lowe (1998, 181) argues that some entities (e.g. surfaces, holes, heaps, events) are

objects, in being countable and in having determinate identity conditions, but are not substances, since incapable of independent existence. More generally, on a conception of substance as capable of independent existence, a Strongly emergent object does *not* count as a substance, since such an object requires the existence of whatever entity or entities are the proper bearers of the base feature. Nomologically as well as conceptually, an emergent entity requires the existence of something else from which it 'emerges'. So there is reason to deny that the new object strategy of response to the collapse objection(s) leads to substance dualism, much less to a problematic form of such dualism.

Finally, it is worth noting that although the new object strategy supposes that fundamentally novel features and powers bring new objects in their wake, it doesn't thereby follow that having Strongly emergent features or novel powers is required for a new object to emerge. Indeed, Weak emergentists also commonly assume (as per Bedau's guiding assumption that an emergent entity is one having an emergent feature) that Weak emergence involves the coming-into-existence of new entities, associated with a distinctive subset of the powers of the base entities/features.

4.3 Objection: compatibility with physical acceptability

I turn now to the concern that satisfaction by a feature *S* of the conditions in *Strong Emergence* is compatible with *S*'s being physically realized, hence physically acceptable.

Yates (2016) offers an argument to this effect. By way of illustrative motivation he argues that the molecular geometry of a water molecule—feature *G*—is a multiply realized feature which bestows powers not bestowed by its realizers. Key to Yates's proposal is the suggestion that some higher-level features are 'qualitatively', as opposed to functionally, realized. Functionally realized features have causal specifications, and correspondingly inherit their powers from their realizers; hence they are at most Weakly emergent. By way of contrast, Yates suggests, qualitatively realized features have noncausal—e.g., mathematical—specifications, which in turn makes room for such features to be causally fundamental, and so Strongly emergent:

> Functionally realized properties […] are at most weakly emergent, but qualitatively realized properties don't share this principled limitation. The reason for this is that the bearer of a qualitatively realized property can have certain causal powers: (i) in virtue of meeting the defining specification, but (ii) not in virtue of basic physical properties and relations in virtue of which it meets that specification. It follows that qualitative realization is consistent with strong emergence […]. (820)

In the case of an H_2O molecule's having molecular geometry *G*, the suggestion is that the specification of *G* is mathematical and spatial rather than causal; that *G* bestows certain powers upon its bearer—in particular, those, including hydrogen bonding in water, associated with the molecule's dipole moment, which is itself explained in terms of *G*; and that (notwithstanding that *G* is physically realized by lower-level physical goings-on) these powers are not had/bestowed by *G*'s realizers.

Granting that the specification of *G* is mathematical and spatial, and that the possession of *G* is associated with various powers, why think that these powers are not had by the base feature *F* that 'qualitatively' realizes *G* on a given occasion? Yates's line of thought seems to be that if such power inheritance were in place, references to *G* could be eliminated in broadly deductive explanations of the dipole moment and associated powers:

> If we attempt to deduce the dipole moment of H_2O from basic physics without appealing to molecular geometry, we draw a blank, for no sooner have we started the deduction than we find ourselves deducing that H_2O has *G* as an essential intermediary step. I conclude that molecular geometry plays a unique role in determining the dipole moment of H_2O, and in virtue of this bestows a conditional power φ that it doesn't inherit from its basic physical realizers. (832)

I respond as follows. First, Yates's methodology is suspect, since it is unclear what metaphysical import should be assigned to considerations about what is required for a given explanation or deduction. Relatedly, nothing in physicalism or in the physicalist supposition that higher-level features inherit their powers from physical base features requires that elements of higher-level explanations, deductive or otherwise, be 'dischargeable' in terms referring only to lower-level physical goings-on. On the contrary, physicalists of all stripes are typically happy to allow that there's no prospect of any such explanatory reductions; relatedly, their reasons for maintaining that higher-level features are at best Weakly emergent (as per their acceptance of *Physical Causal Closure* and the associated strategy for addressing the problem of higher-level causation) do not hinge on the availability of such reductions.

Second, it is especially unclear what import should be assigned to the need to appeal to molecular structure *G* in explanations of the existence and powers of the dipole moment of H_2O. If *G* couldn't be deduced from lower-level physical goings-on, but was nonetheless required in order to metaphysically account for the explananda, that at least would provide some reason for thinking that *G* has powers not had by its dependence base features, along standard (anti-physicalist) Strong emergentist lines; but Yates maintains that *G can* be deduced from lower-level physical goings-on, as an "intermediary step". But then why think that the

need to appeal to *G* indicates that *G* has new powers, as opposed to thinking that this need simply reflects that the explanation of the existence and powers of the dipole moment has to proceed in steps, compatible with the physicalist assumption that any powers of deducible features such as *G* are inherited?[8]

Third (and related to the previous point), Yates observes that his account faces a difficulty:

> At this point, however, a problem arises. I appeal to 'in virtue of' relations twice here: once in the realization relation between the basic physical properties of H_2O and *G*, and again between *G* and conditional powers such as φ, which reflect *G*'s putatively unique role in determining how H_2O molecules are disposed. But if 'in virtue of' is transitive, then H_2O has φ in virtue of its basic physical properties, precluding *G*'s causal fundamentality. (834)

Yates develops this concern as tacitly appealing to a univocal notion of Grounding of the sort I discussed in Ch. 3 (§3.4.4); he then responds that if we rather distinguish two small-'g' grounding relations, then even if there is a kind of abstract Grounding chain here, it is not one entailing that *G*'s powers are inherited:

> The basic physical properties that realize H_2O's molecular geometry are those in virtue of which it meets a mathematical specification; and in virtue of meeting that specification, H_2O has φ. These are two quite distinct grounding relations, but this crucial fact is obscured by focusing on their shared abstract properties. Following Wilson (2014*a*), I distinguish this abstract grounding relation from specific grounding relations such as realization by labelling the former 'Grounding' (with a capital 'g'), and shall argue that Grounding is too coarse-grained to capture the metaphysical structure of my emergentist proposal, which requires two distinct little-g grounding relations. Basic physics mediately grounds φ via one grounding relation—qualitative realization—between basic physics and molecular geometry, and a distinct grounding relation—causal power bestowal—between *G* and φ. That these relations are both instances of Grounding does not entail that φ is really bestowed by the H_2O molecule's basic physical properties.
> . (835)

[8] A related observation attaches to Hendry's (2017) claim that the Strong emergence of molecular structure is motivated by attention to the fact that distinguishing isomers (e.g., ethanol and dimethyl ether) which share the same Schrödinger equation requires "putting in by hand the parameters that specify an important difference between the two cases: the nuclear positions" (153–4). Representational underdetermination aside, presumably the nuclear positions are completely determined by lower-level physical goings-on; but in that case, why think that the need to appeal to such structural details indicates that molecular structure has new powers?

My concern with Yates's response here is that it does not block the natural thought that the relation of qualitative realization is (like functional and other forms of realization) also a relation of causal power bestowal—or, alternatively, that the relation of causal power bestowal is *also* instantiated as holding between G and its physical realizers. For example, suppose some entity has the property of being circular, which specification is given by the usual mathematical formula. Granting that satisfaction of the formula doesn't *in general* entail anything about the having of powers—perhaps there are circular abstracta—in any case whenever concretely realized this geometric feature will be associated with powers. Round cookie cutters have the power to cut round cookies, and so on. But if so, then the concern remains that attention to the grounding chain in Yates's case suggests that G does not have any new powers, after all.

I conclude that the considerations that Yates raises do not provide a compelling basis for thinking that satisfaction of the conditions in *Strong Emergence* is compatible with physical acceptability.

4.4 Objection: non-necessity

I turn next to considering and responding to objections to the claim that satisfaction of the conditions in *Strong Emergence* is necessary for physically unacceptable emergence. There are four main alternative approaches on offer, in terms of epiphenomenalism, supervenience, primitivism, and epistemic criteria, respectively. In what follows, I provide reasons for thinking that each of these alternative approaches to physically unacceptable emergence is unsatisfactory.

4.4.1 Epiphenomenalism

Some attempts to characterize a form of physically unacceptable emergence have aimed to do so in ways that suggest that such emergent features might be epiphenomenal, and so fail not only to have a new power (as per the *New Power Condition*), but to have any powers at all. Hence Morris (2014) says that "we could attempt to characterize varieties of noncausal nonphysicalist novelty that are compatible with supervenience" (353), and Chalmers (1996) endorses a form of what he calls 'naturalistic dualism' which is naturally seen as involving "a limited form of epiphenomenalism" (158–9). Indeed, Chalmers goes on to argue that the sort of 'interactionist dualism' endorsed by the British Emergentists, and which the schema for Strong emergence aims to characterize, doesn't represent any real advantage over epiphenomenalist naturalistic dualism, as regards conscious experience:

> [O]n a close analysis, [interactionist dualism] leaves consciousness as superfluous as before. To see this, note that nothing in the story about emergent causation requires us to invoke *phenomenal* properties anywhere. The entire causal story can be told in terms of links between configurations of physical properties. There will still be a possible world that is physically identical but that lacks consciousness entirely. It follows that at best phenomenal properties *correlate* with causally efficacious configurations. (1996, note 41, 378–9)

The best response to this line of thought, and the associated suggestion that physically unacceptable emergent features might be epiphenomenal, rehearses the response I previously gave in Ch. 3 (§3.3.2) to the concern that satisfaction of the conditions in *Weak Emergence* is compatible with there being noncausal phenomenal aspects of higher-level features. There I argued that it is reasonable to maintain that any phenomenal or qualitative aspects of features would be fully encoded in powers of those features, as per the thesis of Phenomenal Incorporation. There I also argued that the Phenomenal Incorporation thesis is neutral between a physicalist (Weak emergentist) and a nonphysicalist (Strong emergentist) conception of qualitative features. As such, neither the Weak nor Strong emergentist need agree that "nothing in the story about emergent causation requires us to invoke *phenomenal* properties anywhere" such that "The entire causal story can be told in terms of links between configurations of physical properties". On the contrary, either variety of emergentist can reasonably maintain that phenomenal properties are crucially referenced as part of the causal story, and similarly for consciousness more generally (a point to which we will return in Ch. 7).

4.4.2 Supervenience

A common baseline assumption of accounts of emergence, whether of Strong or Weak varieties, is that emergent features minimally nomologically supervene on base features, such that, at least in worlds with the same laws of nature as actually hold, the occurrence of an emergent feature requires the occurrence of some base feature, and (again, in such worlds) any such base feature necessitates the emergent feature. A number of more specific conceptions of supervenience are on offer, aimed at precisifying glosses such as that 'there can be no change in supervenient features without a change in base features' or (equivalently) that 'duplicating the base features duplicates the supervenient features' (see McLaughlin and Bennett 2018 for discussion). For present purposes what is most relevant is, first, that supervenience is an abstract modal correlational notion or relation, typically understood as holding between lower-level and higher-level features, and second,

that some have thought that the difference between physically acceptable and physically unacceptable higher-level features can be cashed simply in terms of a difference in the strength of the correlations holding between higher-level and base features. More specifically, the suggestion is that physically acceptable (Weakly) emergent features supervene with metaphysical necessity on lower-level physical features, such that the correlations are guaranteed to hold in any worlds where the physical features (and, for reasons discussed in Ch. 3, note 15, associated physical laws) are in place, whereas physically unacceptable (Strongly) emergent features supervene with only nomological necessity on lower-level physical features, such that the correlations are guaranteed to hold only in worlds where, in addition to the physical laws, any emergent laws there might be are also in place. For example, van Cleve (1990) characterizes Strong emergence as follows:

> If *P* is a property of *w*, then *P* is emergent iff *P* supervenes with nomological necessity, but not with logical necessity, on the properties of the parts of *w*. (222)

(Here by 'logical necessity', Van Cleve has in mind metaphysical necessity, as characterized above.) Chalmers (2006*a*) endorses a similar conception:

> [W]e can say that Strong emergence requires that high-level truths are not conceptually or metaphysically necessitated by low-level truths. (244, note 1)

The proposed distinction in modal strength here is sometimes taken to serve as a basis for formulating physicalism (understood as per usual as incompatible with Strong emergence) as the thesis that all broadly scientific goings-on supervene with metaphysical necessity on lower-level physical goings-on. Hence Lewtas (2013) says

> The supervenience physicalist [. . .] claims that a suitable physical base metaphysically necessitates [higher-level] properties. But a [Strong] emergent supervenes on its base and yet stands over and above it. Emergent supervenience therefore amounts to lawful correlation and nothing more. (532)

In terms of a global supervenience thesis, the suggestion is that physicalism is true just in case any world duplicating the physical goings-on (including physical laws) duplicates the rest of natural reality:

> Physicalism is true of our world iff any world that is a physical duplicate of our world either is a duplicate of our world *simpliciter* or contains a duplicate of our world as a proper part.[9] (Howell 2009, 85)

[9] The latter condition is intended to accommodate the intuition that worlds that are just like our world except for the addition of ghosts or immaterial souls and the like would not falsify physicalism as holding in this world.

Such supervenience-based characterizations form the basis for an objection to the necessity of *Strong Emergence*, according to which a distinction between strength of modal correlations, rather than a distinction between powers, is what is key to making sense of physically unacceptable emergence.

The supposition that a difference in strength of modal correlation suffices to distinguish physically unacceptable from physically acceptable dependent features is problematic, however. To start, as has frequently been observed, Strongly emergent features might supervene with metaphysical necessity on physical base features (see Horgan 1993, Kim 1998, Levine 2001, Melnyk 2003, Tye 1995, and Wilson 2005). In my (2005), I offer several scenarios illustrating how this might be. One somewhat fanciful but still metaphysically coherent scenario is a version of Malebranchean occasionalism, in which a consistent Malebranchean God brings about certain mental features upon the occasion of certain lower-level physical features, in every world where the latter exist. Another less fanciful scenario involves a view on which features are essentially individuated by (all) the laws of nature into which they directly or indirectly enter (along lines proposed in, e.g, Shoemaker 1980, Swoyer 1982, and Bird 2001, 2002, and 2007). On such a view, any Strongly emergent features there might be would be metaphysically necessitated by base features, since the latter could be instantiated only in worlds containing the emergent as well as the physical laws. Yet another scenario with this consequence is one in which Strongly emergent features involve a nonphysical fundamental interaction, and where it is supposed that the fundamental interactions are unified, in coming as a package deal.

Such scenarios indicate that neither the distinction between metaphysical and nomological necessity, nor the distinction between supervening and not supervening on physical goings-on in worlds where the actual physical laws are operative, can serve to distinguish cotemporally materially dependent features which are physically unacceptable from those that are physically acceptable.[10]

There are two main lines of resistance to this claim; I address each in turn.

The first proceeds by rejecting the seeming counterexamples on grounds that these violate what is sometimes called 'Hume's Dictum', according to which there are no metaphysically necessary connections between wholly distinct existences. As Stoljar (2001) describes the strategy:

> [One] suggestion points out that the problem is only genuine if the cases that generate it are coherent—and are they? One reason against supposing so is that both seem to violate Hume's dictum that there are no necessary connections between distinct existences. According to [Strong] emergentism, for example, mental and physical properties are metaphysically distinct, and yet are necessarily connected.

[10] Note that the concerns here are properly metaphysical. See Kovacs 2019 for critical assessment of broadly epistemological concerns about supervenience according to which it is not an 'explanatory' relation.

Adherence to something like Hume's Dictum is present, for example, in Noordhof's (2010) endorsement of a supervenience-based characterization of the distinction between physically acceptable and physically unacceptable supervenient features:

> The intuitive thought is that the reason why emergent dualists [should] reject appeal to metaphysical necessity is that they suppose that some of the target properties determined by narrowly physical property causes are wholly distinct from them, whereas non-reductive physicalists are committed to thinking that they are not. (71)

A similar thought is operative in Lewtas's (2013) response to certain of the seeming counterexamples, according to which in cases of Strong Emergence, "the correlation connects two entities where [. . .] each is over and above the other" (538).

But appeal to Hume's Dictum is ultimately unsatisfactory by way of response. To start, as Stoljar notes, Hume's Dictum is controversial:

> Hume's dictum is itself a matter of controversy, so it is unclear if the cases can be dismissed in this way (see Jackson 1994, Stalnaker 1996, Stoljar 2010, and Wilson 2005, 2010*c*).

Indeed, post-Humean reasons for believing Hume's Dictum are in short supply. To be sure, if one is a strict empiricist like Hume, who takes the content of our ideas and beliefs to ultimately be a matter of fairly superficial sense experiences, one might well be inclined to endorse Hume's Dictum: any such ideas might, at least in principle, either go together or come apart. But contemporary philosophers are not strict empiricists, and as MacBride (1999) notes,

> [I]t is a curious fact that the proponents of the contemporary Humean programme [...] having abandoned the empiricist theory of thought that underwrites Hume's rejection of necessary connections provide precious little by way of motivation for the view. (127)

Moreover, in a series of papers (see Wilson 2010*a*, 2010*c*, 2014*b*, and 2015*a*) I have considered several potential reasons for accepting Hume's Dictum, and argued that none withstands scrutiny.

Even if Hume's Dictum is accepted, endorsement of this thesis won't sidestep many of the seeming counterexamples, for Strongly emergent features need not be 'wholly distinct' from lower-level physical goings-on (see Wilson 2002*a* and Stoljar 2007). For example, Strongly emergent features might share some (though not all) powers with lower-level physical base features, or be governed by physical as well

as nonphysical interactions. As such, there is no principled route to maintaining a supervenience-based approach to realization that proceeds by way of Hume's Dictum.

A more principled response to defending a supervenience-based approach is suggested by Howell (2009), who argues that the seeming counterexamples to this approach can be rejected, on grounds that if a Strongly emergent feature *S* were to be metaphysically necessitated by a lower-level base feature *P*, then we would have good reason to take *P* to itself be physically unacceptable:

> The basic argument is that if emergence laws are [metaphysically] necessary,[11] and the emergent properties are "new" enough to count as non-physical, then the supervenience base will be polluted and will no longer be purely physical. If this is the case, then [supervenience physicalism] will judge an emergence dualist world to be non-physical, because duplicating the purely physical properties will not duplicate the world simpliciter. (93)

The general suggestion here is that if Strongly emergent feature *S* arises from lower-level feature *P* with metaphysical necessity, then one should take *P* to be individuated, in part, by the disposition to give rise to *S*; and if *P* is so individuated—if part of what it is to be an instance of *P* is to be disposed to give rise to instances of *S*—then one should not regard *P* as a physical property. Howell illustrates his intended point as follows:

> If it turns out that part of what makes electrons what they are is that they give rise to 'unpredictable' qualitative experiences when in a certain setting, then it seems that electrons are somewhat magical and are at least partly constituted by non-physical dispositions. [...] In such a world, a sort of quasi-panpsychism is true: at least some of the basic stuff in our world is not conscious, but it is infused with mentality in that it is individuated by the brute tendency to produce it. (93–4)

In that case, Howell suggests, the possibility of metaphysically necessitated Strongly emergent features poses no threat to characterizing the distinction between Weak and Strong emergence in terms of stronger or weaker modal correlations, since any such features would 'pollute' the supervenience base features in such a way that the latter would no longer be properly considered (lower-level) physical, contra the standard assumption of physicalists and Strong emergentists alike (93).

[11] Note that the antecedent here isn't one expressing that the 'emergence laws' hold in every possible world, but rather that the emergents are connected to their bases in every possible world where those bases exist.

The Strong emergentist has two available responses. The first is due to Morris (2014), according to which "reflection on the base pollution maneuver reveals a new worry about supervenience physicalism" (356). Morris starts by noting that the proponent of a supervenience-based approach to realization (Weak emergence) will, like the Strong emergentist, distinguish the lower-level physical goings-on in the supervenience base from the higher-level features and entities that supervene on this base. In that case, one might naturally wonder whether such higher-level features would also end up 'polluting' the base:

> The base pollution defense supposes [...] that the physicality of a property may be called into question by that property necessarily giving rise to another property that is not itself physical. But in this case, it may seem that supervenient properties will generally end up "polluting the base". Why, in other words, does the problem of "base pollution" specifically concern [Strong] emergentism? Why doesn't it threaten the very idea of properties distinct from [lower-level] physical properties supervening on [lower-level] physical properties? (356)

Howell (2009) is aware of this concern, and addresses it by saying that if the supervenient properties are not "substantively new", then there is no difficulty with maintaining that the base properties are (lower-level) physical (93, note 18). But as Morris (2014) notes, the need to provide an account of when some supervenient features are or are not substantially new introduces a new challenge for supervenience physicalism:

> Granting that the [Strong] emergentist's base cannot be regarded as physical, some account is needed of the conditions under which a supervenient property is not "substantially new" with respect to subvenient, putatively physical properties. Further, without such an account, a supervenience definition of physicalism will provide no guidance for distinguishing between a physical supervenience base and the kind of polluted base associated with [Strong] emergentism, and at least in this way would appear incomplete. The challenge for a supervenience physicalist is to account for the difference between the polluted base and a physical supervenience base without, in effect, rendering talk of supervenience superfluous [if] the only or best way to mark the requisite distinction appeals to the very resources at work in alternative formulations of physicalism or, likewise, alternative accounts of what it is to be a physicalist about some [higher-level] feature of reality. (357)

Morris then goes on to consider two candidate accounts of when a supervenient property is not 'new enough' to result in 'base pollution'—one appealing to a second-order functionalist account, and one appealing explicitly to satisfaction

of the *Proper Subset of Powers Condition*—and to observe that while each account plausibly wards off base pollution, each at the same time delivers a non-supervenience-based account of realization—that is, of physically acceptable emergence—rendering the appeal to supervenience in characterizing such realization otiose. (See also Morris 2018, 45–7.) A similar thought seems appropriately directed at Lewtas's (2013) suggestion that, even if Strong as well as Weak emergents metaphysically supervene on their bases, the latter are distinguished in that "the correlation [is] underlain [. . .] by a more basic constitutive relation" (539).

I agree with Morris's assessment, and moreover add that (as previously argued) insofar as functionally realized features also satisfy the *Proper Subset of Powers Condition*, the more general moral of his assessment is that supervenience-based accounts of realization ultimately do not provide an approach to such realization alternative to one implementing the conditions in *Weak Emergence*.

A second response to Howell's argument is available—namely, to deny his claim that base features will be rendered physically unacceptable if they are disposed to bring about metaphysically necessitated Strongly emergent features or powers. To start, one need not interpret Howell's illustrative case of electrons giving rise to 'unpredictable' qualitative experiences as showing that if base features or entities metaphysically necessitate higher-level features, then 'part of what makes the [base features] what they are' is that they give rise to these higher-level features; nor need one see his case as showing that if base features have dispositions to give rise to physically unacceptable features, the base features will thereby be physically unacceptable. Consider the scenario I previously offered, in which a consistent Malebranchean God brings about mental features on the occasion of certain physical features in every world where the latter exist (that is, with metaphysical necessity). In such a scenario it need not be any part of 'what it is to be' the occasioning physical features that God takes them to be such occasions. Relatedly, one might reasonably maintain that if lower-level physical features do have 'dispositions' to bring about Strongly emergent features, the sense of 'disposition' here is a weak one, reflecting just that the lower-level features are, for whatever reason, metaphysically sufficient preconditions for the Strongly emergent features, which sense implies nothing about the natures of the occasioning features that would impugn their physical acceptability.

That said, on certain proposed counterexamples to a supervenience-based approach to metaphysical emergence, the sense in which a lower-level feature might be disposed to produce a Strongly emergent feature would plausibly inform some part of its nature—for example, if (as previously mentioned; see Wilson 2005 for discussion) features are individuated by all of the laws into which they enter. It remains, even here, that the sense of 'disposition' at issue is weak, again reflecting that the lower-level features are metaphysically sufficient preconditions for the Strongly emergent features. Still, for such cases it seems that more is needed

to block the threat of 'pollution'. And what more is needed seems to be some principled way of characterizing the lower-level features as physical, in spite of their having 'nature-involving' dispositions to produce Strongly emergent features.

One strategy for doing this appeals to the operative characterization of physical entities and features, according to which (as discussed in Ch. 1, §1.4.1) these are the entities and features treated (approximately accurately) by present (and in the limit of inquiry, ideal) physics, and which are not fundamentally mental (see Wilson 2006). Here it is crucial to appreciate two qualifications of the intended characterization of the physical goings-on as satisfying the 'No Fundamental Mentality' (NFM) constraint (or related constraints as discussed in Ch. 1, §1.4.1): first, that the restriction on fundamental mentality is intended to apply to entities and features in comparatively noncomplex combination (what I earlier called 'basic' physical entities and features), reflecting the crucial contrast with panpsychism, according to which individual fundamenta might have or bestow mentality; and second, that the restriction on fundamental mentality pertains not to any unmanifested dispositions there may be, but rather only to manifested features of entities that, again, are comparatively noncomplex. So long as individual lower-level entities and features do not have or bestow mentality, then that they or associated configurations might have dispositions to give rise to Strongly emergent mentality poses no threat to their status as physical or physically acceptable.

Interestingly, while Howell's criterion of physicality imposes a restriction on mentality, his case for taking the restriction to be violated depends on ignoring these qualifications. Howell's argument relies on the following necessary condition on something's being physical:

> ND: Something is physical only if it does not ineliminably involve mental features.

It is the supposed violation of ND by lower-level dispositional features that is supposed to establish that metaphysically necessitated Strongly emergent phenomenal features would undermine the physical acceptability of the dependence base features. But—and notwithstanding that Howell claims that my (2006) characterization of the physical involves ND as a necessary condition—ND is too broad, in failing to specify that the necessary condition here applies only to comparatively noncomplex physical entities and features and moreover only to features manifested under such comparatively noncomplex circumstances. Since the dispositional features under discussion are not so manifest, they pose no threat to the physical acceptability of lower-level features.

A second strategy for characterizing the lower-level features as physical, even given that they have 'nature-involving' dispositions to produce Strongly emergent features, adverts to fundamental interactions. Again, even if, taking all

fundamental interactions into account, lower-level physically acceptable features are essentially disposed to bring about certain Strongly emergent features, it remains that the occurrence of these lower-level features is just a matter of physical fundamental interactions, reflecting the Strong emergentist understanding of lower-level physical features as something like cotemporally sufficient (nomologically or metaphysically, no matter) preconditions for any Strongly emergent features there might be. In short: the Strong emergentist can maintain that the occurrence or instantiation of lower-level physical features ultimately relies just on the operation of fundamental physical interactions, whereas the occurrence or instantiation of a Strongly emergent feature requires the operation of an additional, non-physical fundamental interaction. As in the case of a fundamental interaction-based response to the collapse objection(s), fundamental interactions provide a basis for distinguishing lower-level physical from Strongly emergent goings-on, even when these are deeply dispositionally connected.

4.4.3 Primitivism

I turn next to Barnes's (2012) 'meta-ontological' account of emergence, according to which emergent goings-on are those which are both fundamental and dependent, and where the notions of fundamentality and dependence are each taken to be primitive. Barnes does not distinguish between Strong and Weak forms of metaphysical emergence in her discussion, but in appealing to fundamentality her account plausibly aims to characterize emergence of the Strong, 'over and above' variety.

The general problem with this account is that it is too abstract to satisfy the criteria of appropriate and illuminating accommodation.[12] As we have seen in both historical and contemporary contexts, an understanding of Strong emergence as combining dependence with fundamentality is the traditional starting point of discussions of such emergence, with the bulk of effort devoted to filling in the operative notions of dependence and fundamentality so as to illuminate how these characteristics could be jointly instantiated in ways appropriately accommodating the target cases, while blocking concerns that Strong emergence is naturalistically unacceptable, gives rise to problematic overdetermination, is incoherent (since subject to collapse), and so on. An account of Strong emergence on which the operative notions of dependence and fundamentality are primitive, and where there is no further (or adequate) explication of these notions, goes

[12] See Wilson 2019 for further discussion of concerns with 'abstractionist' accounts of metaphysical emergence, which also include the Grounding-based approaches discussed in Ch. 3 (§3.4.4).

no distance towards illuminating such emergence or addressing the associated concerns.[13]

In particular, though I am friendly to a primitivist account of fundamentality (see Ch. 1, §1.4.3, and Wilson 2014 and in progress)—where better to put primitivity than in what makes it the case that some goings-on are fundamental?—in the absence of further explication about the locus of fundamentality in cases of Strong emergence (e.g., as involving powers that are associated with fundamental interactions), which Barnes does not provide, her account does not ensure that Strong emergents are distinctively efficacious (if, say, the dependent fundamental entity is a noncausal quiddity or other epiphenomenon). A concomitant difficulty is that Barnes's account fails to provide any means of engaging with or addressing either Kim's problem of higher-level causation or the collapse objection.

Another difficulty is that a primitivist account of dependence does not ensure that goings-on that are emergent by lights of Barnes's account are incompatible with goings-on of the sort posited by non-emergentist accounts, such as substance dualism. Barnes does offer an "intuitive gloss" on the notion of dependence, as follows:

> *Ontological Dependence* (OD): An entity *x* is dependent iff for all possible worlds *w* and times *t* at which a duplicate of *x* exists, that duplicate is accompanied by other concrete, contingent entities in *w* at *t*. (880)

But this conception of dependence is too weak to ensure material dependence—again, consider a Malebranchean occasionalist scenario involving a God who consistently (at every possible world) brings about mental states when and only when certain (contingent) physical states are on the scene.[14]

OD also renders Barnes's account subject to a range of counterexamples. Suppose that (as either a Strong or a Weak emergentist might believe) a fundamental physical state *P* necessitates some mental state *M*, distinct from *P*, in any world where *P* exists. Then *P* turns out to be dependent by lights of *OD*, since for all possible worlds *w* and times *t* at which a duplicate of *P* exists, that duplicate will be accompanied by a distinct contingent entity (*M*, or a duplicate thereof). Moreover, since *P* is fundamental, it follows on Barnes's account that *P* is Strongly emergent. That's the wrong result. For similar reasons, Barnes's account interpreted as involving *OD* will deem contingent fundamental determinates, which are

[13] Pearson (2018) and Paolini Paoletti (2018*b*, 14–15) register similar difficulties with Barnes's account.

[14] As Pearson (2018, 396) points out, the conception is also too strong, in the sense that it rules out that Strong emergence might hold with only nomological necessity, as many proponents of such emergence have maintained.

accompanied by their associated determinables in any world where they exist, Strongly emergent. Yet another counterexample is due to Pearson (2018):

> [S]uppose that it were necessary that every time we have charge we have spin and vice versa, and that these were concrete, contingent, Fundamental properties. That wouldn't mean that charge would be an emergent property. Rather, neither charge nor spin appear to be good candidates for emergent properties. (397)

As a supplement to *OD*, Barnes offers a characterization of 'dependence on':

> *x* is dependent on the *y*s iff *x* is dependent because it is part of its intrinsic nature that it bears relation *R* to things intrinsically like the *y*s. (881, note 16)

This is no help, however, since relation *R* is again left uncharacterized, and since the fundamental *x*s at issue might well have (and in the case of determinates, definitely do have) as part of their intrinsic nature that they are intimately related to things intrinsically like the *y*s.

The moral of these multiple difficulties is that there is no shortcut to an adequate characterization of physically unacceptable emergence via primitivist specification of the core notions of fundamentality and dependence. These notions must be filled in, and moreover in a way not leading to immediate counterexample.

4.4.4 Epistemic criteria

A final objection to the necessity claim adverts to satisfaction of one or other epistemic criterion as sufficient for Strong emergence. Indeed, accounts of Strong emergence as involving one or other epistemic failure have been historically common. For example, Broad (1925) characterized emergence as involving in-principle failure of deducibility:

> Put in abstract terms the emergent theory asserts that there are certain wholes, composed (say) of constituents *A*, *B*, and *C* in a relation *R* to each other; that all wholes composed of constituents of the same kind as *A*, *B*, and *C* in relations of the same kind as *R* have certain characteristic properties; that *A*, *B*, and *C* are capable of occurring in other kinds of complex where the relation is not of the same kind as *R*; and that the characteristic properties of the whole $R(A, B, C)$ cannot, even in theory, be deduced from the most complete knowledge of the properties of *A*, *B*, and *C* in isolation or in other wholes which are not of the form $R(A, B, C)$. (61)

And recall Chalmers's (2006*a*) suggestion that

> [W]e can say that Strong emergence requires that high-level truths are not conceptually or metaphysically necessitated by low-level truths. (note 1)

Even if metaphysical necessitation or supervenience, understood as a purely correlational notion, is too weak to characterize physically unacceptable emergence, Chalmers here suggests that a failure of conceptual entailment will do the trick. Finally, consider Horgan's (1993) epistemic account of 'superdupervenience'—that is, supervenience of the sort guaranteeing that supervening properties are nothing over and above their physically acceptable base properties:

> *Horgan's Constraint*: Any genuinely physicalist metaphysics should countenance ontological inter-level supervenience connections only if they are robustly explainable in a physicalistically acceptable way. (563)

Conversely, one might suggest, it suffices for physically unacceptable emergence that a dependent higher-level feature fails to be "robustly explainable in a physicalistically acceptable way".

Though not uncommon, there are two good reasons to reject 'epistemic failure' characterizations of physically unacceptable emergence.

First, these epistemic characterizations are intended by their proponents to track a metaphysical distinction relevant to properly metaphysical emergence. For example, as previously discussed, Broad's epistemic characterization was intended to characterize metaphysical emergence as involving fundamentally novel ('trans-physical') laws and associated fundamentally novel powers. Similarly, the case studies that Chalmers and Horgan offer as illustrating the sort of conceptual entailment or robust explanation that is supposed to be lacking in cases of Strong emergence involve something like functional or causal realization, which are reasonably seen as satisfying the (metaphysical) conditions (in particular, the *Proper Subset of Powers Condition*) in *Weak Emergence*.[15] Conversely, it would

[15] Hence Chalmers (2006*a*) says of heat:

> The concept of heat that we had *a priori*—before the phenomenon was explained—was roughly that of 'the thing that plays this causal role in the actual world.' Once we discover [*a posteriori*] how that causal role is played, we have an explanation of the phenomenon. (45)

And Horgan (1993) says of water:

> Explaining why liquidity supervenes on certain microphysical properties is essentially a matter of explaining why any quantity of stuff with these microphysical properties will exhibit these macro-features [tendency to flow, to assume shape of vessel that contains it, etc.] [...] this suffices to explain the supervenience of liquidity because those macro-features

seem that, on these accounts, the intended sense in which a dependent feature might *fail* to be conceptually entailed or robustly explainable would, as on Broad's account, reflect the feature's having a new power, as per the schema for Strong emergence. As such, the Strong emergentist might maintain that even if Strongly emergent features were always and distinctively (as compared to Weakly emergent features, in particular) accompanied by certain epistemic failures, it would still be advisable to characterize physically unacceptable emergence in terms of what is metaphysically at issue—that is, in the terms encoded in *Strong Emergence*.

Second, in any, case epistemic failures are not distinctive of physically unacceptable emergence. As I will discuss in more detail in Ch. 5: while it made sense in Broad's day to suppose that in-principle failures of deducibility or predictability from lower-level physical goings-on (including laws) were indicative of fundamentally novel laws and associated powers, it has since become clear that many uncontroversially physically acceptable dependent goings-on are not deducible, even 'in principle', from lower-level physical goings-on, for reasons having to do not with fundamental novelty but rather with, e.g., sensitivity to initial conditions (à la the 'butterfly effect') or mathematical limitations of the sort discussed in Boyd 1980, rendering predictions about such goings-on impossible, even given the resources of the entire universe. As such, the proponent of an epistemic characterization of physically unacceptable emergence will need to provide some means of distinguishing unexplained physically unacceptable from unexplained physically acceptable higher-level features.

Here the Strong emergentist can register a variation on the complaint, made by Melnyk (1999) and Morris (2014) against supervenience-based accounts of realization/Weak emergence, that insofar as ensuring the physical acceptability of supervenient entities or features ultimately requires appeal to some or other specific metaphysical relation(s), the appeal to supervenience is rendered 'superfluous' (Morris), 'an idle wheel' (Melnyk).[16] In particular, and assuming satisfaction of the condition on cotemporal material dependence, it is unclear what distinction might be appealed to by way of differentiating physically unacceptable from physically acceptable unexplained phenomena besides one encoding the powers-based conditions in the schemas for Strong and Weak emergence. In that case, however, one can and should dispense with the appeal to lack of explanation (deducibility, conceptual entailment) and rather stick with the powers-based conditions in the schemas, which encode what is really at issue in the contrast between Weak and Strong emergence.

> are definitive of liquidity [and because] it seems explanatorily kosher to assume a "connecting principle" linking the macro-features to liquidity, precisely because those features are definitive; the connecting principle expresses a fact about what liquidity is. (579)

[16] See also Hill 2009, 66 and Morris 2018, 48–9.

4.5 Concluding remarks

In this chapter, I have considered a wide range of objections to the viability of Strong emergence, according to which satisfaction of the conditions in *Strong Emergence* is incompatible with scientific theory or practice, is impossible (as per variations on the 'collapse' objection), is compatible with physical acceptability, or is not necessary for physically unacceptable emergence; and I have argued that each objection admits of one or more responses available to proponents of any of the diverse implementations of the schema. In two cases, additional responses are available which rely on features specific to a fundamental interaction-relative account of such emergence. In particular, such an account provides the basis for distinctive responses to the collapse objection and the associated 'base pollution' objection. My sense is that if any of the considered objections requires tweaking the conditions in *Strong Emergence* (so as to explicitly incorporate reference to fundamental interactions), it is these. If so, the needed relativization to fundamental interactions would be straightforward, and hardly fatal to the overall approach; but again, this is a choice point, and how exactly a Strong emergentist chooses to respond to the various objections may depend on further of their commitments. In any case, these results collectively indicate that *Strong Emergence* is not just a viable and indeed attractively robust means of accommodating physically unacceptable emergence, but that this schema moreover captures what is core and crucial to (and when suitably filled in, what is necessary and sufficient for) such emergence.

5
Complex systems

At this point, I have answered the first key question driving this book: What, more precisely, is metaphysical emergence of the sort seemingly motivated by the special sciences as well as ordinary experience, coupling cotemporal material dependence with ontological and causal autonomy (distinctness and distinctive efficacy)? Underlying the seeming diversity associated with the many accounts of such emergence are two and only two schemas, expressing conditions which are core and crucial to (and when sensibly filled-in, necessary and sufficient for) metaphysical emergence of physically acceptable and physically unacceptable varieties:

> *Weak Emergence*: What it is for token feature *S* to be Weakly metaphysically emergent from token feature *P* on a given occasion is for it to be the case, on that occasion, (i) that *S* cotemporally materially depends on *P*, and (ii) that *S* has a non-empty proper subset of the token powers had by *P*.
>
> *Strong Emergence*: What it is for token feature *S* to be Strongly metaphysically emergent from token feature *P* on a given occasion is for it to be the case, on that occasion, (i) that *S* cotemporally materially depends on *P*, and (ii) that *S* has at least one token power not identical with any token power of *P*.

I have moreover argued that each form of emergence is viable: coherent, metaphysically substantive, naturalistically acceptable, such as to avoid problematic causal overdetermination, and more generally such as to accommodate, in principle, the appearances of seeming metaphysical emergence in an appropriate and illuminating way.

Since there are no in-principle barriers to there being metaphysical emergence of either variety, we are now in position to address (or finish addressing) the second key question driving this book: Is there actually any metaphysical emergence? In this and the following chapters, I will consider this question as applied to four classes of phenomena that have been frequently taken to involve metaphysical emergence: complex systems, ordinary objects, consciousness, and free will.

I begin with complex systems, as perhaps the phenomena that have been most often offered as emergent, by scientists as well as philosophers. Such systems take many forms, both natural (as in cases of turbulent water flows, phase transitions, and weather patterns) and artificial (as in Conway's 'Game of Life', to be discussed in §5.2.1). While there is no agreed-upon definition of what it is to be a complex

Metaphysical Emergence. Jessica M. Wilson, Oxford University Press (2021). © Jessica M. Wilson.
DOI: 10.1093/oso/9780198823742.003.0005

system (see Ladyman *et al.* 2013), the commonly highlighted characteristics of such systems are as follows:

- *Nonlinearity*: Complex systems are nonlinear, in that certain of their features (including associated powers and behaviors) cannot be seen as linear or other broadly additive combinations of features of the system's composing entities, and relatedly, in that mathematical expressions describing the evolution of such systems contain nonlinear terms.
- *Extreme sensitivity to initial conditions*: For 'chaotic' complex systems, small differences in initial conditions can result in huge differences in trajectory or behavior.
- *Unpredictability*: The precise behaviors of complex systems are unpredictable and relatedly, often surprising and seemingly novel.
- *Algorithmic incompressibility*: The dynamical equations of complex systems do not admit of analytic or 'closed' solutions.
- *Universal behavior ('Universality')*: Compositionally different complex systems may exhibit highly similar behavior.
- *Self-organization*: Complex systems exhibit coherent patterns arising as a result of interactions among the parts, in a way suggesting that they are 'self'-organizing.

Such characteristics have been offered as supporting taking certain complex systems to be either Strongly or Weakly emergent; here I present and assess these cases.

In the first section, I consider whether any complex systems might be Strongly emergent. I start with a compressed historical discussion of why the British Emergentists took nonlinearity and in-principle failures of predictability to suffice for fundamental novelty—a view that, while reasonable at the time, was undermined by the discovery and creation of complex systems clearly not involving any fundamentally novel powers/interactions/laws. This discussion is useful for appreciating how nonlinearity moved from being a criterion of Strong emergence to being a criterion of Weak emergence, and for seeing how a recognizable descendant of nonlinearity as a criterion of Strong emergence is present in the aforementioned motivation for new fundamental interactions, reflecting seeming violations of conservation laws. By lights of the latter criterion, I observe, there is little motivation for taking non-mental complex systems to be Strongly emergent—though the case is less clear for certain mental phenomena, a topic to which I return in later chapters. (Those primarily interested in whether complex systems are Weakly emergent can skip this section without loss of continuity.)

In the second section, I consider three existing cases for the Weak emergence of complex systems, appealing to one or other characteristic of such systems, as found in Bedau 1997 and 2008, Mitchell 2012, and Batterman 2000 and 2002. As we'll see, the cases made in these discussions fall short of establishing that complex systems

are Weakly emergent, in failing to rule out certain reductionist strategies for accommodating the characteristics at issue. That said, the prospects for developing these cases in a way that reveals an associated satisfaction of the conditions in *Weak Emergence* are good. In particular, after expanding a bit on my degree-of-freedom (DOF)-based account of Weak emergence, and arguing that attention to order parameters does not undercut its general strategy, I argue that each of the complex systems at issue in Bedau's, Mitchell's, and Batterman's discussions is reasonably seen as Weakly emergent, by lights of a DOF-based account.[1]

5.1 Are complex systems Strongly emergent?

5.1.1 Nonlinearity and unpredictability in the British Emergentist tradition

Perhaps the first suggestion that nonlinearity is indicative of metaphysical emergence is found in Mill 1843/1973, Ch. X, 'On the Composition of Causes'. Recall that Mill distinguishes between two kinds of effects: first, 'homopathic' effects, which conform to "the principle of composition of causes" in being (in some sense) mere sums of the effects of the component causes when acting in relative isolation—as when the weight of two massy objects on a scale is the scalar sum of their individual weights, or when the joint operation of two forces conforms to vector addition in bringing an object to the place it would have occupied had the forces operated sequentially; second, 'heteropathic' effects, which violate the composition principle in not being mere sums in any clear sense. Mill takes this distinction to be crucial, in that (he supposes) the advent of heteropathic—i.e., nonadditive or more generally, nonlinear—effects is indicative of the operation of new laws:

> This difference between the case in which the joint effect of causes is the sum of their separate effects, and the case in which it is heterogeneous to them; between laws which work together without alteration, and laws which, when called upon to work together, cease and give place to others; is one of the fundamental distinctions in nature. (408–9)

By way of illustration, Mill offers chemical compounds and living bodies as entities capable of producing heteropathic effects. Recall also that, given the reciprocal connection between powers and effects, it is straightforward to translate Mill's talk of effects into talk of powers: to say that an effect of a feature of a composite entity is heteropathic, relative to effects of features of its composing parts acting separately, is just to say that the feature of the composite has a power not had by associated

[1] In this chapter I draw on and extend Wilson 2010*b* and 2013*b*.

lower-level base features when in linear combination. Finally, recall that Mill's appeal to the nonlinearity of effects is aimed at identifying a criterion of a higher-level feature's having a fundamentally novel power, enabling it (or its possessing 'body') to override the usual composition laws in the production of certain effects.

Other British Emergentists followed Mill in characterizing Strong emergence in terms of nonlinearity, including Alexander (1920), who also characterized emergent features as having powers to produce heteropathic effects; Morgan (1923), who contrasted emergent features with resultant "additive and subtractive" features; and Broad (1925), who offered scalar and vector addition as paradigms of the compositional principles whose violation was characteristic of emergence. These appeals to nonlinearity, like Mill's, are best seen as attempts to provide substantive metaphysical criteria of the operation of new fundamental powers, forces/interactions, or laws that come into play only under certain complex conditions, as when Broad (1925) says, "[T]he law connecting the properties of silver-chloride with those of silver and of chlorine and with the structure of the compound is, so far as we know, an unique and ultimate law" (64–5). Correspondingly, McLaughlin (1992) accurately characterizes British Emergentism as

> [...] the doctrine that there are fundamental powers to influence motion associated with types of structures of particles [...] In a framework of forces, the view implies that there are what we may call "configurational forces": fundamental forces that can be exerted only by certain types of configurations of particles [...]. (52)

Before continuing, it is worth noting that the operative notion of nonlinearity in the British Emergentist tradition was *not* one according to which it sufficed for nonlinearity (hence, in this tradition, for fundamental novelty) that a feature of a composite entity fail to be a linear combination of *intrinsic* features (powers, etc.) of its composing entities—i.e., a linear combination of features of its composing entities when 'in relative isolation'. Such a contrast would render the view immediately implausible. Consider, for example, the shape of a molecule composed of some atoms. The molecule's shape (and associated powers, etc.) is presumably not the product of any new fundamental interactions; but on the other hand, this shape is clearly not a function (linear or otherwise) just of the shapes of the atoms when in relative isolation—also involved are the spatiotemporal and bonding relations between the atoms associated with these being at some distance from each other.

In light of such cases, the most careful of the British Emergentists—namely, Broad—included pairwise and other relatively noncomplex (e.g., spatiotemporal and lower-level causal) relations between the lower-level entities (or states of affairs consisting of lower-level entities standing in relatively noncomplex lower-level relations) as among the physically acceptable 'summands' apt to be combined in broadly additive fashion, and against which a given claim of emergent nonlinearity

was to be assessed. Hence it was that Broad characterized "pure mechanism" as involving broadly additive deducibility of all higher-level features from features of lower-level entities "either individually or in pairwise combination", and couched his official formulation of emergence in terms of failures of in-principle deducibility of higher-level features from features of lower-level entities both "in isolation" and "in other wholes". The notion of linearity in these construals clearly adverts to lower-level relational as well as intrinsic features of lower-level entities. Such an appropriately broad understanding of linearity can accommodate, e.g., the shape of a molecule, as a (vector) additive function of shapes associated with atoms in pairwise or other relatively noncomplex combination.

Closely related to these appeals to nonlinearity as indicative of fundamental novelty is the British Emergentist supposition that unpredictability, understood to be in some sense 'in principle', is indicative of such novelty. As just noted, Broad's official characterization of emergence was couched in epistemic terms involving in-principle unpredictability:

> The emergent theory asserts that there are certain wholes, composed (say) of constituents *A*, *B*, and *C* in a relation *R* to each other and that the characteristic properties of the whole *R*(*A*, *B*, *C*) cannot, even in theory, be deduced from the most complete knowledge of the properties of *A*, *B*, and *C* in isolation or in other wholes which are not of the form *R*(*A*, *B*, *C*). (1925, 64)

As remarked in Ch. 4, the discussion preceding the formulation makes clear that Broad's appeal to failure of deducibility aims to characterize a metaphysical notion of emergent autonomy, and moreover one tracking a form of emergence incompatible with complete lower-level determination. A similar metaphysical import attaches to Alexander's (1920) claim that insofar as emergents cannot be predicted or explained, they must be accepted with 'natural piety'.

5.1.2 The fall of nonlinearity and unpredictability as guides to fundamental novelty

Given an appropriately sophisticated understanding of the notion of linearity at issue, it was quite reasonable for the British Emergentists to take nonlinearity to be a mark of Strong emergence. To start, various paradigm cases of non-emergent features of composite entities are simple scalar sums of features of their composing entities, as when (bracketing relativistic considerations) the mass of a composite entity is the sum of the masses of its composing entities. More generally and more importantly, at the time it was common to suppose that causal relations ultimately involve the exertion of various fundamental forces,

including gravity and electromagnetism, operating either singly or in tandem. As Mill's discussion makes especially clear, the combination of fundamental forces was taken to proceed in accord with linear composition laws. Again recalling the correspondence between powers and effects, linearity thus looked to provide a general handle on when the features of composite entities did not involve or invoke any new fundamental powers, forces, or laws. Conversely, failures of features or behaviors of composite systems to be subject to linear analysis would have been reasonably interpreted as indicating that some additional fundamental force—a force not operative at lower, less complex levels of natural reality—was now on the scene.

Though reasonable at the time, the British Emergentist supposition that non-linearity is sufficient indication of the fundamental novelty at issue in Strong emergence is no longer plausible. For one thing, the background supposition of causation as involving additive combinations of fundamental push-pull forces is now seen as largely heuristic, or in any case not generalizable; it is fundamental interactions, involving particle exchanges, or yet more abstract accounts of the existence and evolution of natural phenomena, that provide the ultimate story as regards the 'go' of events; and to the extent that Newtonian forces can be seen as real (as they arguably can be; see Wilson 2007) they are now assumed to be nonfundamental (as constituted, somehow or other, by fundamental interactions). For another, in the course of the 20th century, investigations into a wide range of complex systems revealed not just that genuine nonlinearity was rampant, but that in many of these cases the nonlinearity was generated in ways clearly not involving any fundamental novelty. As I noted in my (2013*b*):

> Even as early as the late 1880s, there were difficulties in seeing chaotic complex systems of the sort associated with turbulence in fluids and gasses, and with phase transitions, as linear. Attempts were made to explain away failures of linear prediction in these cases as due to noise or imprecision in measurement; but […] the anomalies and epicycles associated with the supposition of linearity eventually gave way to an understanding of complex systems as being genuinely nonlinear. This is not to say, of course, that failures in prediction were thereby (always) overcome; rather, such failures were given an alternative explanation as reflecting, most saliently, the typical highly sensitive dependence of the associated nonlinear functions on initial conditions (a.k.a. 'the butterfly effect'). (206)

As I further noted, that predictive anomalies in complex systems were not due just to noise or imprecision was confirmed by attention to natural and artificial nonlinear systems for which the relevant initial conditions could be specified with complete accuracy:

> Population growth, for example, is straightforwardly modeled by the nonlinear logistic map:
>
> $$x_n + 1 = ax_n - ax_n^2$$
>
> Here a is a parameter representing birth and death rates, and is different for different systems. The behavior of a given system is heavily dependent on a. For most values of a, the system evolves to a fixed point; as a approaches 4, the system's behavior becomes periodic, and subject to increasingly rapid bifurcation; for $a = 4$, the system's behavior becomes chaotic, with very small differences in initial conditions x_i, associated with distant decimal places, eventually leading to wildly different trajectories. The discovery of natural nonlinear systems encouraged attention to nonlinear systems in general, and with the advent of computers in the latter half of the 20th century, much attention focused on artificial complex systems such as cellular automata, where, as in Conway's 'Game of Life', the stipulated dynamics are nonlinear. (207)

The recognition of genuinely nonlinear systems clearly not involving any additional fundamental forces/interactions or associated novel powers fatally undermined the British Emergentist supposition that nonlinearity is indicative of Strong emergence.

Similar remarks apply to the British Emergentist supposition that unpredictability is so indicative. It was reasonable enough at the time to assume that an insuperable failure of predictability would be an epistemic marker of fundamental novelty: after all, if even a Laplacian demon couldn't deduce certain higher-level phenomena from lower-level goings-on, then what else, besides something fundamentally novel, could explain the occurrence of the higher-level phenomena? But again, increased awareness and appreciation of the distinctive characteristics of complex systems undermined the supposition. In particular, given the extreme sensitivity to initial conditions associated with chaotic complex systems, it turns out to be in-principle impossible to deduce the features of such systems, at least if it counts as a failure of 'in-principle' deducibility that these features could not be deduced even given the resources of the entire universe.

5.1.3 Might some nonlinear phenomena be Strongly emergent?

In light of the previous considerations, it is often taken for granted that complex systems are at best Weakly emergent. Hence Bedau (1997) says:

> An innocent form of emergence—what I call "weak emergence"—is now a commonplace in a thriving interdisciplinary nexus of scientific activity [...] that includes connectionist modeling, nonlinear dynamics (popularly known as "chaos" theory), and artificial life. (375)

As I'll discuss in the next section, there is indeed a good case to be made that many complex systems are (at best) Weakly emergent. Still, it is worth noting that stated reasons for the claim that complex systems are *never* Strongly emergent aren't compelling.

Newman (1996) cites the fact that such systems are 'strictly deterministic' in support of the claim; but nothing prevents Strongly emergent features from entering, both as regards their emergence and their subsequent evolution, into a deterministic nomological net. Nor does the fact that features of complex systems are 'derivable' from nonlinear equations and initial (or boundary) conditions establish physical acceptability, since—as the British Emergentist tradition makes explicit—unlike linear combinations, that nonlinear combinations of physically acceptable features are themselves physically acceptable is not obvious. Bedau (1997) claims that features of complex systems are physically acceptable in being "structural"—that is, in being features of a configuration consisting in some lower-level entities standing in lower-level relations; but given that the features of the configurations at issue do not consist solely in additive combinations of features of their parts, that such features are merely structural (as with, e.g., the shape of a molecule), in a sense that would entail physical acceptability, is again not obvious. One might aim to support the general claim via an argument by analogy, maintaining that insofar as various surprising features of complex systems (period doubling, extreme sensitivity to initial conditions) can be modeled in comparatively simple and artificial systems for which it is uncontroversial that no new fundamental novelty is at issue, there is no reason to suppose that more complex natural complex systems involve fundamental novelty, either. But this argument by analogy fails, for precisely what is at issue is whether, in the more complex natural cases, the nonlinear behaviors at issue have a physically acceptable source.[2]

Indeed, there is in-principle room for maintaining that Strong emergence is at issue in at least some cases of complex systems. Consider, for example, cases where the nonlinear phenomena involve feedback between the micro-entities constituting the base, associated with strange attractors and other dynamic phenomena.

[2] The thought here is not so different from that undercutting the supposition that Strong emergence has been generally discredited by scientific advances; from the fact that some previous candidates for Strong emergence are no longer such candidates, it doesn't follow that all relevant phenomena will admit of treatment in physically acceptable terms.

As Silberstein and McGeever (1999) note, the nonlinearity at issue in complex systems might be taken to involve a kind of system-level holism:

> What is the causal story behind the dynamics of strange attractors, or behind dynamical autonomy? The answer, it seems to us, must be the nonlinearity found in chaotic systems. [...] But why is non-linearity so central? [...] Non-linear relations may be an example of what Teller calls 'relational holism' [...].[3] (197)

Silberstein and McGeever go on to suggest that relational holism of this sort might reflect emergent features' possessing fundamentally new powers ("irreducible causal capacities"), in line with Strong emergence.

It has also been suggested (or interpreted as being suggested) that the singularities standardly associated with thermodynamic phase transitions are indicative of Strong emergence. Hence Menon and Callender (2013) interpret Batterman's claim that "thermodynamics is correct to characterize phase transitions as real physical discontinuities and it is correct to represent them mathematically as singularities" (Batterman 2005, 234) as signaling Batterman's commitment to phase transitions' being emergent along British Emergentist lines. Whether Teller and Batterman would agree that relational holism or thermodynamic singularities should be understood as involving new fundamental powers/interactions/laws is disputable (as I'll shortly discuss in re Batterman's view). Still, for present purposes the crucial point is that one *could* coherently take Strong emergence to underlie some features associated with some complex natural nonlinear systems.

Let's sum up the results so far.

First, whether or not all complex systems are physically acceptable, in any case there's no doubt that some are. The general moral to be drawn from the identification of straightforwardly mechanistic and artificial complex systems is that, contra Mill, Broad, and the other British Emergentists, neither nonlinearity nor (even 'in-principle') unpredictability are sufficient indication of fundamental powers/forces/interactions/laws.

Second, though many complex systems clearly do not involve Strong emergence, stated reasons for the general claim that all complex systems are at best Weakly emergent are uncompelling; as it stands, the general claim is something of an article of faith. It would be nice if, given that nonlinearity and in-principle unpredictability can no longer be seen as criterial of Strong emergence, there were an alternative criterion which could distinguish physically acceptable from physically unacceptable cases of complex systems (assuming any of the latter exist).

[3] In re relational holism, see also the provisional definition of emergence offered by Thompson and Varela (2001).

I turn now to identifying such a criterion, as it enters into my preferred account of Strong emergence.

5.1.4 Nonlinearity's descendant

As previously noted, in Wilson 2002*a* I offer an account of Strong emergence along British Emergentist lines, which fills in the sense in which the power of a Strongly emergent feature is fundamentally novel in terms which explicitly (as opposed to implicitly, as on Broad's characterization in terms of 'in-principle failure of deducibility') involve the coming into play of a new fundamental force/interaction. As above, at the heart of the Strong emergentist position is that some composite systems have features associated with new powers, grounded in new fundamental forces or interactions, suited to enter into producing effects that lower-level configurations can't (directly) enter into producing. McLaughlin (1992) claims that those suspicious of forces (and presumably also of interactions) can dispense with this aspect of the view, retaining only the appeal to new powers. But as we saw in the previous chapter, there is some motivation for thinking that an appeal to fundamental forces or interactions provides perhaps the best response to (among other objections) the 'collapse' objection, according to which the supposedly novel powers associated with Strongly emergent entities or features are always inherited by the base-level entities or features. Recall that on this strategy of response, the collapse objection is avoided by understanding Strong emergence in interaction-relative terms:

> *Interaction-relative Strong Emergence*: Token feature *S* is Strongly emergent from token feature *P* relative to the fundamental interactions in a set {*F*} just in case (i) *S* cotemporally materially depends on *P*, and (ii) *S* has at least one token power that is not identical with any token power of *P* that is grounded only in fundamental interactions in {*F*}.

Again, *Interaction-relative Strong Emergence* makes room for there to be Strong emergence: even if, taking all fundamental interactions into account, any power of a cotemporally materially dependent feature is inherited by its base feature, it remains that higher-level features may be associated with powers that are relevantly 'new', in not being grounded *only* in the set of physical fundamental interactions.

Interaction-relative Strong Emergence also has three features relevant to understanding the bearing of the nonlinearity characteristic of complex systems on Strong emergence.

First, the account nicely accommodates the supposition that features of a composite entity that can be analyzed as broadly additive combinations of physically acceptable features of the composing entities will not be Strongly emergent. For whatever the more specific account of how exactly powers are grounded (schematically speaking) in fundamental interactions, in any case it is clear—to go back to Mill's original discussion—that every power of a feature that is a broadly additive combination of physically acceptable features will be grounded only in fundamental physical interactions, and so fail to be Strongly emergent.

Second, the account suggests an alternative criterion for Strong emergence, which not only survives the advent of physically acceptable nonlinear systems, but moreover provides the means of distinguishing, at least in principle, between cases of nonlinearity that do and cases that don't involve Strong emergence. Here I have in mind the aforementioned criterion of a new fundamental interaction (operative, for example, in the posit of the weak nuclear interaction), adverting to apparent violations in conservation laws. A similar strategy makes in-principle empirical room for testing whether the unusual features associated with complex natural nonlinear systems are or are not due to configurational fundamental interactions, by comparing the values of relevant conserved quantities predicted by fundamental physical theory as attaching to features of composite entities (or associated configurations), with the actually observed values of these quantities. Again: if there's less (or more) energy coming out than going in, for example, we might well be inclined to conclude, following accepted scientific procedure and as per the Strong emergentist thesis, that a new configurational force/interaction has come into play.

Third, in the appeal to apparent violations of conservation laws as a sufficient criterion of Strong emergence we have, it seems to me, a recognizable descendant of the British Emergentist appeal to apparent violations of linearity as such a criterion. For an apparent violation of a conservation law serves, as an apparent violation of a linear composition law was reasonably but incorrectly taken to do, to flag that the whole is more than the mere sum of its parts, such that some fundamentally novel powers (forces, etc.) and laws must be posited, if the sum—of forces, of conserved quantities—is to come out right.

All this said, I take it that there is not much motivation for thinking that any complex natural nonlinear systems involve new fundamental interactions, with the possible exception of those systems associated with qualitative consciousness and free choice (to which we will return in Chs. 7 and 8). Still, it is useful to observe that there is a criterion for Strong emergence upon which (and unlike bare appeals to nonlinearity, relational holism, unpredictability, or representational mismatch) all parties are likely to agree, and which could, in principle, be tested for; and that by lights of this criterion the Strong emergence of at least some complex systems

has not been empirically discredited, since the relevant experiments simply have not been done.

5.2 Are complex systems Weakly emergent?

I turn now to considering whether any complex systems are reasonably taken to be Weakly emergent. My focus will be on three prominent accounts where it is claimed that one or more of the aforementioned characteristics supports seeing complex systems of a certain variety as emergent: first, Bedau's (1997 and 2008) appeal to algorithmic or explanatory incompressibility, applied to properties such as being a glider in Conway's Game of Life; second, Mitchell's (2012) appeal to self-organization, applied to group behaviors such as the flocking of birds; and third, Batterman's (2002 and 2000) appeal to asymptotic universality, as applied to thermodynamic systems undergoing phase transitions.[4] As prefigured at the beginning of this chapter, while the discussions at issue in these accounts do not, as they stand, establish even the Weak emergence of the complex systems at issue, the cases made can be supplemented so as to provide a sound basis for taking complex systems of the target variety to be Weakly emergent.

5.2.1 Bedau's appeal to algorithmic incompressibility

Bedau's focus is on a feature of nonlinear systems shared by both chaotic and nonchaotic nonlinear systems; namely, that such systems typically fail to admit of analytic or closed solutions. The absence of analytic or otherwise 'compressible' means of predicting the evolution of such systems means that the only way to find out what this behavior will be is by going through the motions: set up the system, let it roll, and see what happens. Such algorithmic incompressibility serves as the basis for Bedau's (1997) account of physically acceptable emergence, applicable in contexts where a composed system *S* has "microstates" (encoding intrinsic states of its parts) whose time evolution is governed by a microdynamic *D*, and where *S*'s "macrostates" are structural properties constituted wholly out of its microstates:

[4] I do not discuss Newman's (1996) treatment of the feature *being in the basin of a strange attractor*, since his treatment (according to which such a feature is identical to some lower-level feature, albeit in such a way that it is 'epistemically impossible' for us to discover which one) is explicitly ontologically reductive. Newman's discussion is representative of approaches on which the emergence of complex systems is purely epistemological (see also Popper and Eccles 1977, Klee 1984, and Rueger 2000). However illuminating such accounts may be as regards why we find the features and behavior of such systems interesting, novel, or unpredictable, they do not provide any basis for taking complex systems to be even Weakly metaphysically emergent. The accounts I will consider in the text aim, or can be seen as aiming, to provide such a basis.

> Macrostate P of S with microdynamic D is *weakly emergent* iff P can be derived from D and S's external conditions but only by simulation. (378)

Derivation of a system's macrostate "by simulation" involves iterating the system's microdynamic, taking initial and any relevant external conditions as input. The broadly equivalent conception in Bedau's (2002) takes physically acceptable emergence to involve 'explanatory incompressibility', where there is no shortcut explanation of certain macrostates of a composite system. In being derivable by simulation from a microphysical dynamic, such macrostates are plausibly physically acceptable, such that emergence understood as involving incompressibility "is consistent with reasonable forms of materialism" (Bedau 1997, 376).

By way of illustration, Bedau focuses on Conway's Game of Life, an example of a nonchaotic nonlinear map. The game consists in a set of simple rules, applied simultaneously and repeatedly to every cell, each of which may be 'alive' or 'dead', in a two-dimensional lattice. At each step in time, cells are updated as per the following rules: (1) any live cell with fewer than two live neighbors dies; (2) any live cell with two or three live neighbors stays alive; (3) any live cell with more than three live neighbors dies; (4) any dead cell with exactly three live neighbors becomes alive. Here there is no problem of sensitivity to initial conditions, since these conditions consist just in the discrete 'seeding' of the lattice. Still, Bedau argues that the property of being a glider in the Game of Life is emergent, in being algorithmically incompressible. That this property does not involve any Strong emergence is clear, since for cellular automata the long-term behavior of the system is completely metaphysically determined by ('derived from') the lower-level rules applying to cells in the grid. But that a given system will evolve in such a way as to generate a glider can typically not be predicted from knowledge of initial conditions (seeding) and these rules.

Before continuing, it is worth registering that, notwithstanding that Bedau describes the failure of predictability at issue as being diachronic (involving an inability to antecedently predict the later evolution of a given system), he aims to motivate a view of the relevant complex systems as emerging from "underlying processes" (1997, 395), and (as I'll shortly discuss) as manifesting macro-level patterns. Correspondingly, we can think of Bedau's account as characterizing a form of emergence in which certain temporally extended macrostates (or collections thereof) cotemporally depend on underlying microstates (or collections thereof), in ways that due to algorithmic incompressibility could not be antecedently predicted.

Now, Bedau's appeal to algorithmic incompressibility, like Newman's appeal to epistemically inaccessible identity (see note 4 of this chapter), might be thought to support complex systems' being merely epistemically emergent, an impression seemingly confirmed when Bedau says that "weakly emergent phenomena are

autonomous in the sense that they can be derived only in a certain non-trivial way" (2002, 6). Indeed, Bedau explicitly claims that emergent features of composite systems are both ontologically and causally reducible to features of their composing systems:

> [W]eakly emergent phenomena are ontologically dependent on and reducible to micro phenomena. (2002, 6)
>
> [T]he macro is ontologically and causally reducible to the micro in principle. (2008, 445)

Notwithstanding these reductive claims, Bedau maintains that algorithmic incompressibility provides a basis for genuinely metaphysical emergence. He offers two reasons for thinking this, but (as I discuss in Wilson 2015*b*), neither establishes the point.

The first is that the incompressibility of an algorithm or explanation is an objective metaphysical (if broadly formal) fact:

> The modal terms in this definition are metaphysical, not epistemological. For *P* to be weakly emergent, what matters is that there is a derivation of *P* from *D* and *S*'s external conditions and any such derivation is a simulation. [...] Underivability without simulation is a purely formal notion concerning the existence and nonexistence of certain kinds of derivations of macrostates from a system's underlying dynamic. (1997, 379)

But such facts about explanatory incompressibility, though objective and hence in some broad sense 'metaphysical', are not suited to ground, in particular, the ontological or causal autonomy of emergent entities. What is needed for such autonomy is not just some or other metaphysical distinction between macro- and micro- goings-on, but moreover one which plausibly serves as a basis for rendering the higher-level features at issue ontologically and causally autonomous—that is, distinct and distinctively efficacious—with respect to the lower-level features upon which they cotemporally materially depend.

The second reason Bedau gives is more promising—namely, that the algorithmically incompressible features of complex systems typically enter into macro-level patterns and laws. As Bedau says:

> [T]here is a clear sense in which the behaviors of weak emergent phenomena are autonomous with respect to the underlying processes. The sciences of complexity are discovering simple, general macro-level patterns and laws involving weak

> emergent phenomena. [...] In general, we can formulate and investigate the basic principles of weak emergent phenomena only by empirically observing them at the macro-level. In this sense, then, weakly emergent phenomena have an autonomous life at the macro-level. (1997, 395)

As such, Bedau maintains, "weak emergence is not just in the mind; it is real and objective in nature" (2008, 444). Attention to macro-level patterns is a move in the right direction towards gaining emergent autonomy; but, two points. First, I don't see how Bedau can maintain that some emergent goings-on are not "merely epistemological" (2008, 451) and rather reflect an "autonomous and irreducible macro-level ontology" (2002, 38) , while also maintaining that these same goings-on are ontologically and causally reducible to micro-level goings-on. Either the ontological/causal autonomy or the ontological/causal reducibility has to go.

Second, for purposes of blocking the potential reducibility of higher-level to lower-level features, it isn't enough merely to point to the fact that the higher-level feature enters into macro-level patterns (and associated laws) which are in some sense more general than the patterns (and associated laws) into which the lower-level features enter. For the reductionist has various strategies for accommodating such general patterns (by, e.g., taking the feature whose having corresponds to instantiating a given pattern to be identical with a disjunction of specific lower-level ways of implementing the pattern). Hence while something seems right about attending to the comparative generality of the patterns associated with features like *being a glider*, more needs to be said if reductionist accommodation of this sort of pattern is to be blocked. Bedau does not say more along these lines, however, and the upshot is that this strategy, like the previous one, does not in itself establish that any complex systems are even Weakly metaphysically emergent.

That said, as prefigured I will later argue that there are good reasons to think that a DOF-based account can accommodate the Weak emergence of the sort of features that Bedau discusses. Moreover, it seems likely that the sorts of arguments that I gave in Ch. 2 (§2.3) supporting the ontological and causal autonomy of realized features on various accounts of realization (including functionalist, mechanist, mereological, determinable-based, and 'ontological explanation' accounts) could be brought to bear here; for these accounts also typically aim to realistically accommodate macro-patterns and general laws, and (as I argued) are reasonably seen as blocking reductionism via satisfaction of the conditions in *Weak Emergence*. Bedau's account of physically acceptable emergence might well be profitably understood as combining one or more of these accounts of realization with the further stipulation, unique to certain nonlinear complex systems, that the behavior of these systems is algorithmically or explanatorily incompressible.

5.2.2 Mitchell's appeal to self-organization

I next turn to Mitchell's (2012) account of the emergence of chaotic complex systems exhibiting what is sometimes called 'dynamic self-organization'. In general, Mitchell follows Kauffman (1993, 1995), Wimsatt (1994, 1996, 2007), Bedau (1997, 2008), Thompson and Varela (2001), Thompson (2007), Camazine *et al.* (2018), and others in taking emergence to involve "certain types of non-aggregative compositional structures" (179). Mitchell moreover maintains that in the case of certain chaotic complex systems, the non-aggregativity is dynamic, in arising in a process-like fashion from interactions of the constituents, and in involving feedback loops of the sort characteristic of self-organized systems:

> Self-organized systems are ones in which feedback interactions among simple behaviors of individual components of a system produce what appears to be an organized group-level effect. (183)

The flocking of birds is a case in point:

> Simple additive relations and simple linear equations [...] will fail to make sense out of much of the complexity that we find in nature even though patterns and structures emerge from the simple interactions of the constituents. The vee pattern that emerges in a flock of geese or the more complex patterns of flocking starlings are not predictable by an aggregation of behaviors of individuals in solo flight, but only from the non-aggregative interaction or self-organizing that derives from the local rules of motion plus feedback among the individuals in group flight (see Couzin 2007; see Rosen 2007 for photos of the starling patterns). Ontologically, there are just physical birds; there is no new substance, no director at a higher level choreographing the artistic patterns the flocks make. Nevertheless, this type of behavior is emergent. (179)

Here again, the characterization of the purported emergence in dynamic and diachronic terms can be reconceived as involving the emergence of certain macro-level patterns (e.g., a 'vee' pattern) from the spatiotemporally distributed lower-level configurations of individuals upon which the patterns cotemporally materially depend.

Mitchell is sensitive to the threat of ontological and causal reduction faced by certain accounts of physically acceptable emergence, but maintains that features arising from dynamic self-organization are not subject to such reduction, on grounds, first, that "interaction among the parts generates properties which none of the individual components possess" and second, that "these higher-order properties in turn can have causal efficacy, i.e., novelty" (179). I address each of these motivations for blocking reduction, in turn.

First, as previously discussed (e.g., in Ch. 1, §1.4.2), a conception of emergence according to which it suffices for the operative notion of 'non-aggregativity' that a composed whole have "properties which none of the individual components possess" is too weak to serve as an appropriate dialectical basis for this notion. Again, the reductionist will happily allow that configurations (pluralities or structural aggregates) often and indeed typically have properties that are 'novel' in the weak sense of not being had, either by way of type or token, by individual components. Hence it is that if there is to be an interesting distinction between reductionists and emergentists of whatever stripe, the reductionist must be granted resources suited unto making sense of (massively complex, spatiotemporally extended) lower-level configurations and associated features.

Moreover, these additional resources need not be restricted to linear combinations, contra what Mitchell seems to suggest in observing that "simple additive relations and simple linear equations" do not suffice to accommodate flocking behavior. For—not least because fundamental physics is itself nonlinear—the reductionist is within their rights to take any lower-level interactions (linear or nonlinear, as the lawful case may be) between components into account as input into their theorizing—at least when those interactions are manifest when in relatively non-complex combinations.[5] But in that case, it is unclear why the reductionist cannot agree with Mitchell that there is "scientific emergence" as involving "concrete accounts of how and why rules of interaction among components produce difficult-to-predict emergent behaviors". The upshot is that neither the fact that certain properties of flocks of birds are not had by individual birds, nor the fact that these group-level properties are nonlinear functions of properties of individuals, blocks the ontological reduction of the group-level features at issue.

Nor does Mitchell's discussion of the causal novelty associated with group-level behaviors block causal reduction. Mitchell notes a number of ways in which group-level dynamics is associated with interesting causal interactions. These include that self-organization results in stable properties (pertaining, e.g., to the density of flocks being concentrated at the interior) which can be "the target of natural selection [and hence] exhibit causal saliency (see Carere *et al.* 2009)" (180); that behaviors giving rise to group-level properties can also, if heritable, explain certain facts about natural selection; and that (as per experiments run by Page and Mitchell on 'computer bees', discussed in their 1991 and 1998) groups can come to manifest division of labor, even when the individuals are assigned very 'minimal' behaviors. Mitchell (2012) suggests that these forms of causal interaction are distinctive in incorporating historical and contextual features:

[5] This is one place where a law-based conception of the individuation of levels appears to be superior to an 'ontologically lightweight' conception. At a minimum, the latter approach needs to be somehow modified to allow for some nonlinear combinations of features at a level to be placed at that level.

> It is still the case that the properties and behavior of the component parts cause the ensuing behavior of the system, but there is a shift of emphasis to the features of the history and context that the system experiences to understand why one outcome occurred rather than another. (182)

But here again, the reductionist will presumably maintain that, while these forms of causal interaction are massively complex, there is nothing—at least for all Mitchell's discussion of the cases establishes—to prevent their being given a reductionist treatment. In particular: nothing requires that causal interactions on a reductionist account be spatiotemporally local or somehow free of history or context; on the contrary (and as a causal variation on the theme of my Ch. 3, §3.3.3 response to Melnyk), the reductionist can allow that spatiotemporally 'wide' circumstances crucially enter into what causes what. It is moreover worth noting that the reductionist will point to Mitchell's reference to 'a shift in emphasis' from local to broader features as suggestive of, and in any case as consonant with, a merely pragmatic reading of the higher-level efficacy at issue. So far, then, Mitchell's stated considerations do not establish that self-organizing behaviors are anything more than highly complex lower-level processes.

As prefigured, I will later argue that a DOF-based account can accommodate the Weak emergence of the sort of features that Mitchell discusses. Moreover, it is again worth noting that the considerations that Mitchell raises, pertaining to group-level effects and stable patterns of behavior, are consonant with the considerations highlighted in the various accounts of realization discussed in Ch. 2 (§2.3), and it is correspondingly likely that Mitchell could avail herself of certain associated strategies for achieving ontological and causal autonomy. In particular, it is plausible that, whatever exactly 'self-organization' comes to, in any given case this will be associated with a functional role that can be implemented by diverse (collections of) configurations involving lower-level interactions between lower-level components. In the case of flocking behavior, for example, the functional role will advert to the instantiation of a vee pattern, which pattern and associated role can be implemented by diverse configurations of locally interacting birds. Hence it is natural to see the features to which Mitchell is calling attention as a variation on the theme of functional realization, on which the functional role at issue has the distinctive characteristics of self-organizing systems.

5.2.3 Batterman's appeal to asymptotic singularities

Batterman has written a great deal concerning the status as emergent or reducible of special-science entities (see his 2002, 2005, and elsewhere), though there has remained unclarity as regards whether he takes emergence to be a metaphysical

rather than merely epistemic or representational phenomenon, and if so, of what strength. As I'll shortly discuss, there is reason to think that Batterman does not have even a Weak account of metaphysical emergence in mind, much less the Strong account some attribute to him. Still, Batterman's work on emergence in asymptotic regions of the sort associated with phase transitions is relevant to the present discussion, both because chaotic nonlinear systems are associated with such transitions, and because, whatever Batterman's view of the matter, at least one of the features of systems in asymptotic regions that he mentions can, as I'll later discuss, support taking complex systems to be Weakly metaphysically emergent.

Let's start with the basics of Batterman's account of asymptotic emergence. An asymptote in mathematics is a limiting value of a function that is approached indefinitely closely, but never reached. For example, as $x \to 0$ the function $1/x$ goes to infinity; in this (though not in every) case the asymptote is associated with a discontinuity. Interestingly, many 'near-neighbor' scientific theories involve asymptotes: special relativity asymptotically approaches Newtonian mechanics in the limit as $v/c \to 0$, wave optics asymptotically approaches ray optics as $1/\lambda \to 0$, quantum mechanics asymptotically approaches Newtonian mechanics as Planck's constant approaches 0, and statistical mechanics asymptotically approaches thermodynamics in the 'thermodynamic limit', where particle number N and volume $V \to$ infinity. Now, in the latter three cases, the asymptotic limits at issue are associated with discontinuities in the regions near the asymptote. In such cases of 'singular' asymptotic limits, Batterman suggests, we have reason to take various objects or features associated with the asymptotic region (or associated theory) to be emergent. In particular, Batterman suggests that various features of systems undergoing phase transitions, including those associated with certain critical exponents, are emergent features of such systems. Why so, and in what sense?

Batterman's most explicitly stated reasons for thinking that phase transitions involve emergence cite certain explanatory factors. One explanatory concern reflects a kind of theoretical mismatch between the near-neighbor theories at issue, insofar as the discontinuities associated with taking the thermodynamic limit, and which are commonly supposed to be needed to accommodate the associated asymptotic phenomena, find no representational mirror in the analytic functions of statistical mechanics. Even if there were no problem with deriving specific instances of asymptotic features from the micro-theory, however, a second explanatory factor would remain—namely, that the characteristic universality of asymptotic phenomena cannot be properly explained by reference just to lower-level 'causal-mechanical' explanations. As above, the behavior of systems undergoing phase transitions is characterized by critical exponents. As Batterman (1998) says,

> What is truly remarkable about these numbers is their universality […] the critical behavior of systems whose components and interactions are radically different is virtually identical. Hence, such behavior must be largely independent of the details of the microstructures of the various systems. This is known in the literature as the "universality of critical phenomena". Surely one would like to account for this universality. (198)

Lower-level causal-mechanical explanations, even if available, cannot account for universality, for, as Hooker (2004) puts it, these "will be infinitely various in detail and this will block any reconstruction of what is universal about them" (442).

By way of contrast, Batterman argues, various methods for modeling asymptotic phenomena—most notably, the Renormalization Group (RG) method—do provide an explanation of the universal features of systems undergoing phase transitions. The RG method takes a system's governing laws (e.g., its Hamiltonian) and iteratively transforms these into laws having a similar form but (reflecting moves to increasingly 'larger' scales) fewer parameters—i.e., DOF. In the limit of applications of the method, the resulting Hamiltonian describes the behavior of a single 'block', corresponding to the macroscopic system:

> [T]he transformation effects a reduction in the number of […] degrees of freedom within the correlation length. […] In effect, the renormalization group transformation eliminates those degrees of freedom (those microscopic details) which are inessential or irrelevant for characterizing the system's dominant behavior at criticality. (200)

That the method is suitably applied to systems undergoing phase transitions reflects that near critical points, such systems cease to have any characteristic length scale, and are 'self-similar' in that the laws governing the systems take the same form at all length scales.[6] Application of the RG method to such systems enables calculation of the critical exponents associated with phase transitions, and so provides an explanation of the universal behavior of systems near critical points:

> [I]f the initial Hamiltonian describes a system at criticality, then each renormalized Hamiltonian must also be at criticality. The sequence of Hamiltonians thus generated defines a trajectory in the abstract space that, in the limit as the

[6] See also Lamb 2015: "In terms of phase spaces, RG methods […] systematically [reduce] the dimensionality of infinitely or nearly infinitely dimensional systems with the result that certain behaviors of distinct complex systems represented by distinct high-dimensional phase spaces may be represented by an identical low-dimensional phase space. After renormalization, only the dimensions required to account for the behavior of interest are included in this new, 'renormalized', phase space" (90–1).

> number of transformations goes to infinity, ends at a fixed point. The behavior of trajectories in the neighborhood of the fixed point can be determined by an analysis of the stability properties of the fixed point. This analysis also allows for the calculation of the critical exponents characterizing the critical behavior of the system. It turns out that different physical Hamiltonians can flow to the same fixed point. Thus, their critical behaviors are characterized by the same critical exponents. This is the essence of the explanation for the universality of critical behavior: Hamiltonians describing different physical systems fall into the basin of attraction of the same renormalization group fixed point. […] This stability under perturbation demonstrates that certain facts about the microconstituents of the systems are individually largely irrelevant for the systems' behaviors at criticality. (201)

Batterman's account of asymptotic emergence cites three features potentially relevant to the emergence of many complex systems:

1. elimination of micro-level degrees of freedom (DOF)
2. universality of certain features or behavior
3. stability of certain behavior under perturbation.

As Hooker (2004) notes, these features are generally characteristic of chaotic nonlinear systems:

> In every case of so-called 'critical phenomena' […] the asymptotic domain shows a universally self-similar spectrum of fluctuations. […] This is indicative of chaos and occurs when behaviours are super-complexly, but still systematically, interrelated. (440)

Indeed, the core similarities between critical phenomena in statistical mechanics and chaotic nonlinear phenomena, including period doubling and intermittency routes to chaos of the sort displayed by the logistic map, have led to an active area of investigation in which "the logistic map is [understood as] a prototypical system […] for the assessment of the validity and understanding of the reasons for applicability of the nonextensive generalization of […] Boltzmann-Gibbs statistical mechanics" (Mayoral and Robledo 1970, 339). If one or more of the features entering into Batterman's account can be seen as supporting metaphysical emergence, this result would apply to a wide range of complex systems.

Now, as it happens (and as I discuss further in Wilson 2013*b*) there is good reason to believe that Batterman does not intend to be seen as offering either an account of or a case of metaphysical emergence. As above, he is rather focused on the question of what is required if the critical behaviors of the systems in question

are to be explained. Hence Morrison (2012) correctly reads Batterman as offering an 'explanatory' account of emergence:

> I characterize Batterman's account [of emergence] as explanatory insofar as the main argument centers on how asymptotic methods (via the RG) allow us to explain features of universal phenomena that are not explainable using either intertheoretic reduction or traditional causal mechanical accounts [...]. (143)

That said, nothing prevents us from considering whether any of the aforementioned features—elimination in micro-level DOF, universality, and/or stability under perturbation—might serve as sufficient indicators of metaphysical emergence. With an eye to sticking somewhat closely to Batterman's work, we might look especially to universality and stability under perturbation, since it is these features which he has most consistently highlighted as motivating the 'emergence' not just of systems near critical points, but also of special-science entities to which the RG approach and its associated strategy for eliminating DOF do not apply.

Universality and stability under perturbation are really two sides of the same coin; as Batterman (2000) says, "most broadly construed, universality concerns similarities in the behavior of diverse systems" (120). The suggestion that physically acceptable emergence might be a matter of universality or of stability under micro-level perturbations is common enough; indeed, we saw something similar in Bedau's (1997) conjecture that the fact that certain features of nonlinear automata (e.g., gliders) enter into 'macro-level patterns' might support such features' being ontologically autonomous. It is unsurprising, then, that the same concerns with Bedau's conjecture also attach to attempts to locate metaphysical emergence in universal or stable features of composite entities—namely, that reductionists have various strategies for accommodating such features. What is needed, if these features are to be seen as tracking the ontological and causal autonomy of composite entities, is a better response to the usual reductivist strategies; but Batterman does not provide such a response—unsurprisingly, given that his primary concern is with whether an appropriately explanatory account of universality is available, not with whether such features are ontologically or causally autonomous.

That said, it is likely that attention to the powers associated with features manifesting universality or stability under perturbation would provide a basis for establishing such autonomy. In particular, these characteristics are clear variations on the theme of multiple realizability; and as I argued in Ch. 2, there are cases to be made, on various accounts of realization, that multiply realizable features have only a proper subset of the token powers of their realizers, at both the level of types and the level of tokens, and so (in combination with the cotemporal dependence

condition, which is clearly satisfied in cases of complex systems undergoing phase transitions) satisfy the conditions in *Weak Emergence*.

In the remainder of this chapter, I want to focus on the third feature Batterman discusses—namely, that of a system's having degrees of freedom (DOF) that are eliminated with respect to the system of base entities upon which it cotemporally materially depends. In the next section, I'll present a DOF-based account of Weak emergence, as developed in Wilson 2010*b*, and argue that it is applicable not only to complex systems undergoing phase transitions, but also to other complex systems, including Bedau's gliders and Mitchell's flocking birds.

5.2.4 A DOF-based account of Weak emergence

Call states upon which the law-governed properties and behavior of an entity *E* (object, system, or other particular) functionally depends the 'characteristic states' of *E*. A DOF is then, roughly, a parameter in a minimal set needed to describe an entity as being in a characteristic state. Given an entity and characteristic state, the associated DOF are relativized to choice of coordinates, reflecting that different sets of parameters may be used to describe an entity as being in the state. More precisely, the operative notion of DOF is as follows:

> *Degrees of Freedom (DOF)*: For an entity *E*, characteristic state *S*, and set of coordinates *C*, the associated DOF are parameters in a minimal set, expressed in coordinates *C*, needed to characterize *E* as being in *S*.

I'll sometimes speak for short of 'characterizing an entity', with state and coordinates assumed.

Some common characteristic states, and DOF needed to characterize certain entities as being in those states, are as follows:

- *The configuration state*: tracks position. Specifying this state for a free point particle requires 3 parameters (e.g., x, y, and z; or r, ρ, and θ); hence a free point particle has 3 configuration DOF, and a system of N free point particles has $3N$ configuration DOF.
- *The kinematic state*: tracks velocities (or momenta). Specifying this state for a free point particle requires 6 parameters: one for each configuration coordinate, and one for the velocity along that coordinate; hence a free point particle has 6 kinematic DOF, and a system of N free point particles has 6N kinematic DOF.
- *The dynamic state*: tracks energies determining the motion. Specifying this state typically requires at least one dynamic DOF per configuration

coordinate, tracking the kinetic energy associated with each position coordinate; other dynamic DOF may track internal/external contributions to the potential energy.

Why attend to DOF in hopes of illuminating Weak emergence? To start, as above, different entities, treated by different sciences, may be functionally dependent on the same characteristic state (e.g., the configuration state). Moreover, as above, the DOF needed to characterize intuitively different entities as being in these states typically vary. Following these observations, I take the main cash value of attention to DOF to lie in the fact that DOF track the details of an entity's functional dependence on its characteristic states, in a more fine-grained way than the mere fact that the system is in the state does. Driving my account is the idea that the fine-grained details concerning functional dependence that are encoded in the DOF needed to characterize broadly scientific entities serve as a plausible ontological basis for the individuation of such entities.

I start by observing an important tripartite distinction (again, see Wilson 2010*b*) relevant to such functional dependence, reflecting that the DOF needed to characterize an entity may be reduced, restricted, or eliminated in certain circumstances (typically associated with the imposition of certain constraints or more generally, the presence of certain energetic interactions), compared to those needed to characterize a (possibly distinct) entity, when such circumstances are not in place. To prefigure: eliminations in DOF, in particular, enter into the DOF-based account of Weak emergence. Let's get acquainted with these different relations and note an example of each.

First, constraints may *reduce* the number of DOF needed to characterize an entity as being in a given state. So, for example, a point particle constrained to move in a plane has 2 configuration DOF, rather than the 3 configuration DOF needed to characterize a free point particle. In cases where a particular DOF is given a fixed value, the laws governing an entity so constrained are still functionally dependent on the (now constant) value of the DOF; hence such constraints do not eliminate the DOF, but rather reduce it to a constant value. By way of example: rigid bodies treated in classical mechanics have such a reduced set of DOF relative to the unconstrained system of their composing entities.

Second, constraints may *restrict* the DOF needed to characterize an entity. A point particle may be constrained, not to the plane, but to some region including and above the plane. Characterizing such a particle still requires 3 configuration DOF, but the values of one of these DOF will be restricted to only certain of the values needed to characterize the unconstrained particle. Cases of restriction in DOF are more like cases of reduction than cases of elimination of DOF, in that, again, the entity's governing laws remain functionally dependent on specific values

of the DOF. By way of example: molecules, whose bonds are like springs, have a restricted set of DOF relative to the unconstrained system of their composing entities.

Third, constraints may *eliminate* the DOF needed to characterize an entity. For example, N free point particles, having $3N$ configuration DOF, might come to compose an entity whose properties and behavior can be characterized using fewer configuration DOF, not because certain of the DOF needed to characterize the unconstrained system are given a fixed value or restricted to a range of values, but because the properties and behavior of the composed entity are functionally independent of these DOF. A case in point is that of a spherical conductor of the sort treated in electrostatics, which has DOF that are eliminated relative to the system of its composing entities; for while the E-field due to free particles depends on all charged particles, the E-field due to a spherical conductor depends only on the charges of particles on its surface. Certain quantum DOF are also eliminated in the classical (macroscopic) limit. For example, entities of the sort treated by classical mechanics are ultimately composed of quantum entities, but the characteristic states of classical-mechanical entities do not functionally depend on the spins of their quantum components.

Two features of such special-science case studies are commonly assumed, and I will follow the presentation in my (2010*b*) in taking them for granted in what follows as holding of the special-science entities under discussion. First is that the holding of the constraints relevant to reducing, restricting, or eliminating DOF occurs as a matter entirely of physical or physically acceptable processes. Such processes suffice to explain why sufficiently proximate atoms form certain atomic bonds, why atoms or molecules engage in the energetic interactions associated with SM ensembles, and so on. More generally, for each of the aforementioned special-science entities E, the constraints on the e_i which enter into their composing E are explicable using resources of the theory treating the e_i (or resources of some more fundamental theory, treating the constituents of the e_i). Call such constraints 'e_i-level constraints'. A second important feature of the case studies is that it is again plausible that all of the properties and behavior of the special-science entities E at issue are completely determined by the properties and behavior of their composing e_i, when these stand in the relations relevant to their composing E.

In my (2010*b*), I drew on the case studies and associated features as motivating the following (here, renamed) DOF-based account of Weak emergence:

DOF-based Weak Emergence: An entity E is Weakly emergent from some entities e_i if

1. E is composed by the e_i, as a result of imposing some constraint(s) on the e_i.

2. For some characteristic state S of E: at least one of the DOF required to characterize the system of unconstrained e_i as being in S is eliminated from the DOF required to characterize E as being in S.
3. For every characteristic state S of E: every reduction, restriction, or elimination in the DOF needed to characterize E as being in S is associated with e_i-level constraints.
4. The law-governed features of E are completely determined by the law-governed features of the e_i, when the e_i stand in the relations relevant to their composing E.

The conditions should be understood as relativized to occasions, as per usual. Here the first and fourth conditions encode the supposition of cotemporal material dependence of an emergent entity E and its features; on the usual supposition that the e_i are physically acceptable, the first condition more specifically encodes physical monism and the fourth condition more specifically encodes *Physical Causal Closure*. The second condition, against the backdrop assumption of the fourth condition, encodes the satisfaction of the *Proper Subset of Powers Condition* for at least one feature (characteristic state) of E, for reasons that I will present shortly. Finally, the third condition encodes the assumption that any constraints associated with features of E (which constraints, as discussed in Ch. 3, are typically operative in cases where a feature satisfies the *Proper Subset of Powers Condition*) are a result of only e_i-level (i.e., physically acceptable) processes.

One last point before we proceed: in this account I cut out the feature 'middleman' and focus directly on the emergence of an entity E. If we want to express a DOF-based account of physically acceptable emergence in terms of features, this is perhaps most straightforwardly done by replacing reference to E with reference to the token feature of being an entity of type E, and similarly for the configuration of e_i upon which E cotemporally materially depends. Here I'll stick with the entity-based formulation, for continuity with discussion in my previous article.

The ontological irreducibility of DOF-eliminated entities

Systems that are emergent by lights of the above account are physically acceptable, given that the unconstrained system of composing entities is physically acceptable. In my (2010*b*), I argue for this in some detail, but here I'll just observe that this result is to be expected, given that both the constraints relevant to composing E, as well as all of E's law-governed features (hence associated powers), are physically acceptable (in addition to, as above, satisfying the cotemporal dependence condition). More crucial is to establish that entities satisfying the conditions in *DOF-based Weak Emergence* are ontologically and causally autonomous.

I start by expanding on the general DOF-based strategy for accommodating the ontological and causal autonomy of Weakly emergent goings-on. Let us suppose

that some entities e_i are treated by the lower-level physical laws L_p, and that an entity E satisfies the conditions in *DOF-based Weak Emergence* with respect to the e_i, such that (among other things) the DOF needed to characterize E as being in some characteristic state are strictly fewer than those needed to characterize the system of lower-level physical e_i as being in that state. Suppose also, to fix ideas and as is typically the case for macro-entities in the classical limit (a point to which I will return in Ch. 6, §6.1.1), that among the DOF eliminated in characterizing E are those pertaining to quantum spin. In this case, E will not be appropriately identified with any lower-level physical goings-on, whatever they might be. Why not? Because (as previously noted in Ch. 3, §3.2.3) the goings-on appropriately placed at L_p will be those whose specification includes all the DOF needed for the physical laws to operate, including quantum spin. Since E's characterization, by assumption, fails to include (DOF containing) any information about quantum spin, the lower-level physical laws will not be able to take an entity such as E as input. E, and more generally any goings-on whose DOF are eliminated vis-à-vis their lower-level physical realizers, cannot be identified with any lower-level goings-on, since their specifications fail to include all the DOF needed for the lower-level physical laws to operate. This much shows that an entity E's satisfaction of the conditions in *DOF-based Weak emergence* suffices to ensure that E is ontologically autonomous—that is, distinct—from any lower-level goings-on.

Relatedly, attention to the relation between the DOF characterizing E and the DOF characterizing its realizing system(s) also provides a basis for E's satisfaction of the *Proper Subset of Powers Condition*, and so for E's being causally autonomous—that is, distinctively efficacious—with respect to E's realizers. Indeed, there are two results to this result.

The first route proceeds by attention to the bearing on E's powers of E's having strictly fewer DOF than the realizing system of composing e_i. To start, as per the fourth condition in *DOF-based Weak Emergence*, the law-governed features of E on a given occasion are completely determined by the law-governed features of the system of e_i on that occasion. On the intended understanding of 'complete determination', it follows that every token power of E, on a given occasion, is identical to some token power of the system of e_i, such that E satisfies the *Token Identity of Powers Condition*. The reverse is not the case, however: for the system of e_i, but not E, has token powers associated with quantum spin. It follows that the system of e_i has more token powers than E, and the *Proper Subset of Powers Condition* in the schema for Weak emergence is satisfied. Correspondingly, and as previously discussed, considerations of difference-making and/or comparatively abstract systems of laws provide a sound basis for taking E to be distinctively efficacious with respect to the system of e_i.

The second route proceeds by attention to the bearing on *E*'s powers of its being treated by a special science that is 'special', in particular, in treating entities such as *E* only under certain conditions which are in some or other sense prerequisites for the eliminations in DOF at issue. Given that (as we are assuming, for purposes of illustration) *E* is a composed entity whose satisfaction of the conditions in *DOF-based Weak Emergence* involves elimination of DOF associated with quantum spin, the relevant conditions would be those pertaining to the existence of composed entities in the classical limit, including that temperatures and energies be low enough to allow for composed entities to exist. This provides a second basis for satisfaction of the *Proper Subset of Powers Condition*. As I put it in my (2015*b*):

> [W]hat powers an entity has are plausibly a matter of what it can do; and the sciences are plausibly in the business of expressing what the entities they treat can do. It follows that, plausibly, what powers an entity has are expressed by the laws in the science treating it. The powers of *E* are thus those expressed by the laws in the theory treating (constrained) entity *E*, while the powers of [the system of e_i] are those expressed by the laws in the more fundamental theory treating the (relatively unconstrained) lower-level constituents of [the system of e_i]—that is, the constituents of [the system of e_i] as existing both inside and outside the constraints associated with *E*. Consequently, the laws of the theory treating *E* express what happens when certain lower-level entities stand in relations associated with certain lower-level constraints, and the laws treating [the system of e_i] express what happens when certain lower-level entities stand both in these relations and in other relations not associated with the constraints. Hence [the system of e_i] has more powers than *E*, and the proper subset relation between powers in Weak emergence is thus in place. (387)

Order parameters and DOF

Before going on to apply a DOF-based account of Weak emergence to the case of complex systems, I want to address the potential import of a related feature of certain complex systems having eliminated DOF—namely, the association of such systems with an order parameter. Both Morrison (2012) and Lamb (2015) agree that elimination of DOF is characteristic of Weak emergence in complex systems, but they also each discuss order parameters as equally and perhaps more crucial. For example, Morrison says:

> The basis of the idea of universality is that the fixed points are a property of transformations that are not particularly sensitive to the original Hamiltonian. What the fixed points do is determine the kinds of cooperative behavior that are possible, with each type defining a universality class. The important issue here is not just the elimination of irrelevant degrees of freedom; rather it is the existence

> or emergence of cooperative behavior and the nature of the order parameter (associated with symmetry breaking) that characterizes the different kinds of systems. (160)

Lamb (2015) suggests something similar, albeit pitched in terms of explanation:

> A reductive explanation is one in which all of the constraints are characterized in terms of small-scale DOF. [...] [R]epurposing some of Wilson's terminology, for an explanation to be reductive it must be the case that the DOF needed to explain why E exhibits S includes only DOF that characterize the smallest scale DOF of e_i [...]. When a system's components have minimal or linear coupling it is very likely that an adequate explanation of the system's behaviors and features will be reductive. A non-reductive explanation is one for which an order parameter is identified in the explanans. In this case, the order parameter identifies large-scale features or behaviors that constrain the phenomena of interest. In terms of DOF, non-reductive explanation is an explanation for which at least one of the DOF used to explain E as being in $[C]$ is not a DOF required to characterize the smallest-scale $[e_i]$ in the set $[e_i]$ exhibiting σ. (71)
>
> The order parameter identifies a large-scale constraint on relatively smaller features of the system, such that the behaviors of the smaller features are dependent on system-wide organizational features. (74)

One might correspondingly wonder if the real indicator of metaphysical emergence in the cases at hand is reflected in the introduction of an order parameter rather than an elimination in lower-level DOF. And one might moreover wonder whether, as Lamb seems to suggest, such a parameter might be seen as introducing a new DOF, associated with a new power to constrain lower-level goings-on, contra the assumption, in *DOF-based Weak Emergence*, that operative constraints on the lower-level entities are themselves lower-level. Supposing so, it might be thought that attention to the DOF associated with complex systems near critical points suggests that such systems are Strongly, not Weakly emergent. As Lamb (2015) notes, "If the order parameter characterization identifies a novel physical feature or power, not characterizable in terms of a smaller scale, then that is our best evidence for strong emergence" (117).

Lamb goes on, however, to register that he suspects that "strong emergence of this form would not be unscientific or a violation of physicalism" (117). Similarly, Morrison (2012) is clear that, for the cases of symmetry breaking (as in the transition to a ferromagnetic or superconducting state) that are her primary focus, the underlying micro-physics is completely determinative of the higher-level phenomena:

> Symmetry breaking is reflected in the behavior of an order parameter that describes both the nature and magnitude of a broken symmetry. In the ferromagnetic state, the order parameter is represented by the vector describing the orientation and size of the material's magnetization and the resulting field. In the superconducting case, the order parameter is the amplitude φ of the macroscopic ground state wave function of the Cooper pairs. The electromagnetic properties in a superconductor are dominated by Cooper pairs, whereas electrons in a metal normally behave as free particles that are repelled from other electrons because of a negative charge. Because Cooper pairs appear only at T (their presence indicates that the system has undergone a phase transition), they give rise to the order parameter, which involves the Cooper pairs forming a single wave function. In general, the order parameter can be thought of as an extra variable required to specify the macroscopic state of a system after the occurrence of a phase transition. In nonsuperconducting metals, gauge invariance ensures that $\varphi = 0$. (152)

Three observations are relevant at this point. First, notwithstanding talk of phase transitions' (or symmetry breakings') 'introducing' or 'giving rise to' order parameters, both Lamb and Morrison register that the parameter is present at the micro-level. Indeed, standard discussions of order parameters take these to be present at the micro-level, so to speak: what is new upon the occurrence of a phase transition is not a given order parameter itself, but rather the *value* of the parameter, such that (in a simple toy case) the parameter for a metal prior to magnetization has the value zero, and the same parameter after magnetization has the value one. More realistically, that a given order parameter is present prior to a given phase transition is required in order for it to be able to register the range of values associated with the coming-to-be of the ordered state, as when more and more electrons become 'Cooper pairs' on their way to composing a superconductor. But in that case, the parameter, or associated DOF, isn't in fact 'new'.

Second, it is worth noting that on one common understanding of order parameters, they are not properly seen as DOF, understood as parameters needed in order to specify the law-governed properties and behavior of some characteristic state of a given entity (object or system). Rather, order parameters are seen as mere 'phenomenological descriptions' of the extent to which an entity manifests a certain variety of order. As Kleman and Lavrentovich (2003) put it:

> The concept of an order parameter has appeared with the attempt to describe the order-disorder transition of alloys, specifically, to define a degree of disorder [...] First elaborated by Gorsky and Bragg and Williams to describe order-disorder transitions in alloys, it has been developed in its modern form by Landau for the purpose of a phenomenological description of phase transitions. (76)

This much suggests that, pace Morrison, it is the eliminations in DOF rather than the 'introduced' order parameter that are most important in understanding the sense in which complex systems undergoing phase transitions are emergent.

Third, it is worth recalling that metaphysical emergence involves cotemporal material dependence of emergents on base goings-on. As such, in considering whether phase transitions involve genuine metaphysical emergence, it isn't obviously to the point to consider whether a given feature associated with a phase transition (say, superconductivity) is metaphysically emergent from some states of affairs prior to the phase transition. The concern here, as in cases of purported diachronic intra-level emergence (as in Humphrey's cases of fusion, or 'transformative' emergence, of which particle annihilation is the primary case in point) is that what is being offered as a case of emergence is better seen as simply an interesting case of diachronic causation (see Ch. 1, note 7).

For the foregoing reasons, it is reasonable to assume that attention to order parameters as involved in complex systems' undergoing phase transitions poses no threat either to the physical acceptability of such systems or to the potential applicability to such systems of an account of emergence in terms of eliminations in DOF.

5.2.5 *DOF-based Weak Emergence* and complex systems

Let's now return to the question of whether any complex systems are plausibly seen as Weakly emergent. As I'll now argue, there are cases to be made that many such systems, including those having the characteristics discussed by Bedau, Mitchell, and Batterman, are Weakly emergent by lights of a DOF-based account.

I start with Batterman's account of complex systems undergoing phase transitions, having the features of universality and stability under perturbation. Given Batterman's observation (discussed in §5.2.3) that the Renormalization Group (RG) method applies to such complex systems, and given that eliminations in DOF (along with certain other suppositions which are here in place, concerning e_i-level constraints and e_i-level determination) are sufficient for Weak emergence, we can be comparatively brief.

As previously noted, the RG method applies to systems undergoing phase transitions, which are relevantly similar to and indeed can be understood as chaotic nonlinear systems; and that the RG method applies to a given system provides as good an indication as we are likely to get that the system has DOF that are not just reduced or restricted, but eliminated as compared to the unconstrained system of its composing entities. We can thus argue as follows:

1. Systems that can be modeled by the RG method have eliminated DOF (Batterman 1998 and elsewhere).
2. Chaotic nonlinear systems are modeled by the RG method.
3. Therefore, chaotic nonlinear systems have eliminated DOF.
4. Systems with eliminated DOF (and which satisfy other conditions satisfied by chaotic nonlinear systems) are Weakly emergent (Wilson 2010*b*).

∴ Chaotic nonlinear systems are Weakly emergent.

As confirmation of the fact that many chaotic systems have eliminated DOF, it is worth noting that one of the puzzles that Batterman raises for thermodynamic systems carries over to chaotic complex systems, and is answered in just the same way. The puzzle he raises concerns how thermodynamic systems can be a viable object of study. Such systems—e.g., an isolated gas E—are composed of massively large numbers of particles or molecules e_i. Since (boundary restrictions aside) the composite entity E in this case is effectively unstructured, shouldn't it have the same DOF as the system of unconstrained e_i? Supposing so, however, the success of statistical mechanics (SM) is mysterious, since obviously we are not in position to track such massive numbers (on the order of 10^{26}) of DOF. As Batterman (1998) puts it:

> One wants to know why the method of equilibrium SM—the Gibbs' phase averaging method—is so broadly applicable; why, that is, do systems governed by completely different forces and composed of completely different types of molecules succumb to the same method for the calculation of their equilibrium properties? (185)

The answer reflects that while the e_i are not bonded, they are interacting via exchanges of energy, and such interactions may not only restrict or reduce, but eliminate DOF, as indicated by the RG method's being appropriately applied to such systems. This, then, is the answer to the puzzle: such systems are tractable since the modes of interaction of their composing entities result in their having DOF that are massively eliminated compared to the unconstrained system of composing entities. Again:

> [T]he renormalization group transformation eliminates those degrees of freedom (those microscopic details) which are inessential or irrelevant for characterizing the system's dominant behavior at criticality. (Batterman 1998, 200)

A similar puzzle applies to chaotic complex systems. Recall that chaotic complex systems are characterized by their extreme sensitivity to initial conditions. If nonlinear systems are so sensitive and their resulting trajectories so "chaotic", how

is it that they can be, as they are, a viable object of scientific study? The answer, I take it, is effectively the same: the composing entities, though not bonded, are energetically interacting, in ways that, as application of the RG method reveals, massively eliminate DOF needed to characterize the composite system. Here we have a solution to the puzzle, and more to the present point, a decisive, empirically supported case for taking the important class of chaotic nonlinear systems to be Weakly metaphysically emergent.

Though applicability of the RG method to a given complex system provides a quick route to taking the system to be Weakly emergent, such applicability is not necessary in order to establish that a given complex system satisfies the conditions of Weak emergence (DOF). Indeed, there are cases to be made that the complex systems targeted in Bedau's and Mitchell's discussions also satisfy the conditions of this account, and so are appropriately taken to be Weakly emergent.

First, recall Bedau's suggestion that certain phenomena, such as gliders in Conway's Game of Life, manifest a sort of autonomy characteristic of 'innocent'—that is, physically acceptable—emergence, as involving behaviors that are "autonomous with respect to the underlying processes [involving] simple, general macro-level patterns and laws" (1997, 395). In this artificial case, we do not have laws of nature on hand to look to as a guide to the DOF of the entities at issue. Nonetheless, the rules of the Game of Life will do for these purposes.

In particular, we can consider what DOF are required to specify the location of the system of 'live' cells composing a 'macro-entity' in the Game of Life as compared to the DOF required to specify the location of the macro-entity itself. To focus on a simple case, consider a glider (the composed entity E) composed of five 'live' cells, which starts at the origin in the configuration shown in Figure 5.1:

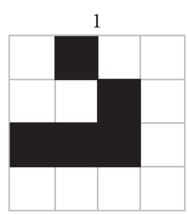

Figure 5.1 Beginning state of the glider

The glider 'moves' in a four-step sequence, at the end of which it returns to its original configuration, one grid diagonally to the SE, so to speak, as shown in Figure 5.2:

Now, let us suppose that the initial seeding of the grid contains only the initial state of a glider, so that we can ignore any potential collisions of the glider with

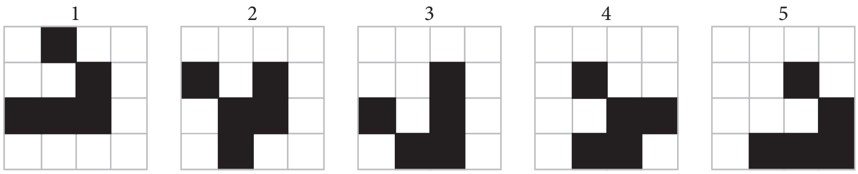

Figure 5.2 Four-step periodic evolution of the glider

other 'entities'. Given that the Game of Life takes place on a (potentially infinite) two-dimensional grid, the location of each of the cells composing the glider E will require two DOF, corresponding to values (relative to some arbitrarily chosen origin point) along the x and y axes; hence determining the position of each of the five 'live' cells composing a glider, in ways that can then be input into the 'lower-level' rules of the Game of Life, will require ten DOF. By way of contrast, the position of a glider at n time-steps from t_0 requires only two DOF, specifying the x, y coordinates of the relevant diagonal cell, which at a time-step $n/4$ will be associated by the 'glider laws' with the relevant stage of the glider. As such, a glider E in the Game of Life has DOF that are eliminated relative to the system of its composing e_i.

Moreover, a glider E clearly satisfies the other conditions at issue in *DOF-based Weak Emergence*. It is composed by the e_i (the cells of the grid), as a result of imposing some constraint(s) on the e_i, where the constraints here simply reflect how the rules of the Game of Life introduce a stable configuration. There is a characteristic state S—namely, position—the specification of which for E requires strictly fewer DOF than those required to characterize the system of e_i. Every reduction, restriction, or elimination in the DOF needed to characterize E as being in a characteristic state is associated with e_i-level constraints, since any such constraints are a matter only of the operation of the rules of the Game of Life. And more generally, the law-governed features of E are completely governed by the law-governed features of the e_i.

Correspondingly, the considerations previously brought to bear showing that an entity E satisfying the conditions of *DOF-based Weak Emergence* with respect to its system of composing e_i can here be brought to bear in support of Bedau's claim that gliders in the Game of Life are Weakly emergent.

Next, recall Mitchell's suggestion that certain complex phenomena, such as flocking birds, manifest dynamic self-organization. Here the composing entities are in the first instance the individual birds making up the flock. Specifying the position of each bird requires at least three DOF (actually, it would require many more, but for present purposes this won't matter); hence specifying the position of 100 birds would require 300 position DOF. When the birds compose a flock, however, they mutually constrain each others' positions, moving (to idealize

somewhat) as an ensemble. Such constraints, as noted, can introduce restrictions, reductions, or eliminations in the DOF needed to specify the position of the birds. Now, since a flock isn't a rigid body, and since knowing the position of the flock at a time requires knowing its shape at that time, the DOF needed to specify the position are all those needed to specify this shape. For this it suffices to know the position of every bird on the external boundary of the flock; the positions of birds in the interior of the flock are irrelevant.[7] Since some birds in a flock will be in the interior, it follows that the DOF needed to specify, at a time, the position of a flock, will be strictly fewer than—will be eliminated as compared to—the DOF needed to specify, at a time, the positions of all the individual birds constituting the flock.

That said, the case of flocking birds, unlike the case of the other complex systems we have considered, introduces a new consideration as entering into the system at issue—namely, consciousness, at least of a perceptual variety. Presumably it is some part of birds' flocking the way they do that they perceive and are otherwise aware of other birds in the flock and what they are doing, and adjust their own behavior accordingly. Of course individual birds are also perceptually conscious, so this fact doesn't in itself suggest that flocks fail to satisfy the conditions in *DOF-based Weak Emergence*. But if the entities composing the flock are further decomposed into lower-level physical entities, then whether flocks are Weakly emergent from lower-level physical entities will hinge on the status as either Strongly or Weakly emergent of such conscious mental features—a topic to which we will return in Ch. 7.[8]

5.3 Concluding remarks

Complex systems have frequently been offered as cases of actual metaphysical emergence, but as I've argued in this chapter, previous accounts of such systems as either Strongly or Weakly emergent do not, as they stand, establish this result.

This is commonly acknowledged as regards British Emergentist claims that nonlinearity and unpredictability are sufficient marks of fundamental novelty of the sort associated with Strong emergence—though as I'll discuss in Chs. 7 and 8, the empirical jury is still out over whether complex systems associated with mentality manifest apparent violations in conservation laws, which criterion is,

[7] How exactly to determine this boundary is a good question. Perhaps some sort of density measure would come into play. In any case, the point in the main text doesn't hinge on this epistemological issue.

[8] This point highlights another way (beyond the relativization to fundamental interactions associated with my preferred account of Strong emergence) in which attributions of emergence may be relative. For all that has been established (or ruled out) thus far, it might be that flocks are Weakly emergent from birds, but birds (*qua* conscious entities) are Strongly emergent from lower-level physical phenomena.

I've here suggested, a recognizable descendant of nonlinearity as a marker of fundamental novelty.

By way of contrast, it has frequently been claimed that certain characteristic features of complex systems, including, in addition to unpredictability and nonlinearity, the manifesting of macro-patterns, self-organization, universality, and stability under perturbation, themselves serve as a sufficient basis for taking the complex systems at issue to be Weakly emergent. As I've argued here, however, such claims require supplementation, to the extent that there are available reductionist strategies for accommodating these features that need to be blocked if the claim of emergence is to be defended.

It may be—indeed, it is likely—that such supplementation can be achieved by attention to the sorts of considerations canvassed in the Ch. 2 discussion of functionalist and other accounts of realization. In any case, there is another feature of complex systems—namely, having DOF some of which are eliminated as compared to the DOF needed to characterize the associated realizing system—which provides a principled basis for blocking the threat of ontological and causal reduction, in a way compatible with physicalism, given the physical acceptability of the realizing system. This feature is most clearly present in cases of complex systems undergoing phase transitions, to which the RG method (which effectively works by eliminating DOF) applies. But as I've also argued, there are cases to be made that certain macro-patterns in cellular automata (e.g., gliders) and certain biological complex systems (e.g., flocks of birds) also have eliminated DOF, and more generally satisfy the conditions in *DOF-based Weak Emergence*.

The upshot is that we now have confirmation of the common but previously unsubstantiated belief that at least some complex systems are emergent in a way compatible with physicalism—that is, are Weakly emergent.

6
Ordinary objects

In this chapter, I turn to the question of whether ordinary objects are either Strongly or Weakly metaphysically emergent. By 'ordinary' objects I have in mind objects which are uncontroversially inanimate (as Thomasson, 2007 puts it) or nonliving (as Merricks, 2003 puts it), and of the sort with which creatures like us are or may be perceptually acquainted.[1] Such objects might be either natural (rocks, feathers, mountains, planets) or artifactual (tables, baseballs, statues).[2] While there are several competing metaphysical accounts of the nature of objects,[3] the discussion to follow will be broadly neutral on which metaphysical account of objects is correct, so long as this account does not rule out of court the possibility that ordinary objects are metaphysically emergent.

Here the order of consideration is reversed from that of the previous chapter. In the first section, I consider whether any ordinary objects are either Weakly emergent or (as I will sometimes put it) are 'at least' Weakly emergent, in having at least one feature satisfying the conditions in the schema for Weak emergence.[4] First, I argue that ordinary objects of the sort appropriately treated by classical (or 'Newtonian') mechanics are Weakly emergent by lights of a DOF-based account, for reasons anticipated in the previous chapter; second, I argue that a common conception of artifacts as associated with sortal properties and distinctive functional roles supports thinking of these as being at least Weakly emergent by lights of a functional realization account; third, I argue that ordinary objects typically have metaphysically indeterminate boundaries, which when coupled with an attractive determinable-based account of such indeterminacy, indicates that such ordinary

[1] In its focus on inanimate ordinary objects, the purview of the present chapter is more restricted than that at issue in Korman 2020, according to which "ordinary objects are objects belonging to kinds that we are naturally inclined to regard as having instances on the basis of our perceptual experiences: dog, tree, table, and so forth".

[2] Some (e.g., Grandy, 2007) suggest that there is no deep difference between artifactual and natural objects; here I'll assume that there is a distinction, though as we'll see, much of my discussion of the one type of object applies, mutatis mutandis, to the other.

[3] Among the many approaches are those on which objects are bundles of properties (either tropes, as per Campbell 1990, or universals, as per Paul 2002), combinations of substrate and property (as per Locke 1690 and Simons 1994), and hylomorphic compounds (as per Fine 2003 and Koslicki 2008).

[4] Recall (as prefigured in Ch. 1, §1.2.3) that an object's being 'at least' Weakly emergent is sufficient for its not being reducible to lower-level goings-on, but leaves open the possibility that some other of its features satisfies the conditions in the schema for Strong emergence, in which case the object would be deemed Strongly, not Weakly, emergent.

Metaphysical Emergence. Jessica M. Wilson, Oxford University Press (2021). © Jessica M. Wilson.
DOI: 10.1093/oso/9780198823742.003.0006

objects are at least Weakly emergent, by lights of a determinable-based account of such emergence.

In the second section, I consider whether any ordinary objects are Strongly emergent. Here I argue that the best case for this stems from artifactual ordinary objects whose functional or other characterization reflects or encodes certain social practices involving normative or aesthetic goings-on. The ultimate status of such objects as Strongly or rather just Weakly emergent hinges, like the status of certain complex systems involving mentality, on the status as Weakly or Strongly emergent of the associated mental features of persons, of the sort to be discussed in the next chapters.

I close by observing that the results of this chapter serve not just to establish the status of ordinary objects as at least Weakly emergent, but also to undercut the motivations for Thomasson's meta-ontological view, as discussed in her (2010) and elsewhere, according to which investigations into the ontological status of artifactual ordinary objects should proceed differently from investigations into the ontological status of special-science entities.

6.1 Are ordinary objects Weakly emergent?

6.1.1 Classical objects

What I will call 'classical' objects are ordinary objects of the sort whose static and dynamic behaviors are appropriately treated by classical or Newtonian mechanics, understood as comprising, roughly, Newton's three laws of motion and the gravitational and electromagnetic force laws. Classical objects include special-science entities such as rocks, feathers, planets, and other comparatively structurally stable inanimate objects. Artifacts may also count as classical ordinary objects; but to avoid introducing complications associated with mentality (to be discussed later in this chapter) I will restrict my focus here to special-science objects.

Classical mechanics as a special science

To start, it's worth saying a few words in support of taking classical mechanics to be a special science. It is sometimes suggested that classical mechanics is at best a false but useful approximation to more fundamental theories. Hence Cohen-Tannoudji *et al.* (1977) say

> Classical laws cease to be valid for material bodies [...] on an atomic or subatomic scale (quantum domain). However, it is important to note that classical physics [...] can be seen as an approximation of the new theories, an approximation which is valid for most phenomena on an everyday scale. (9)

Though natural, this line of thought can and should be resisted, at least if it is taken to suggest that classical mechanics is irrelevant to understanding the metaphysical structure of natural reality.

To start, it is characteristic of special sciences that they have restricted application: the laws of thermodynamics, chemistry, cell biology, geology, botany, neurobiology, psychology, and so on, traversing the ladder of theories of constitutional and biological complexity, do not hold in every context, but rather require for their appropriate application that certain constraints or boundary conditions be in place (as requisite for, e.g., the occurrence of thermodynamic systems, complex inorganic and organic structures, specific genetic or environmental factors, and so on). Hence while the restricted application of classical mechanics, along with its presumably derivative status with respect to fundamental physics, spells its failure as a fundamental theory, it remains a candidate for being a *nonfundamental* theory, with something important to tell us about the nature of nonfundamental reality. Indeed, like other special sciences, classical mechanics is reasonably seen as tracking an important and distinctive level of ontological grain, and associated laws of nature, salient in circumstances wherein lower-level details are irrelevant to the dynamics of (relatively) large-scale goings-on. This is just how Feynman (1965) characterizes classical mechanics:

> Newton's laws are the "tail end" of the atomic laws, extrapolated to a very large size. The actual laws of motion of particles on a fine scale are very peculiar, but if we take large numbers of them and compound them, they approximate, but *only* approximate, Newton's laws. Newton's laws then permit us to go on to a higher and higher scale, and it still seems to be the same law. In fact, it becomes more and more accurate as the scale gets larger and larger. [...] As we apply quantum mechanics to larger and larger things, the laws about the behavior of many atoms together do *not* reproduce themselves, but produce *new laws*, which are Newton's laws, which then continue to reproduce themselves from, say, micro-microgram size, which still is billions and billions of atoms, on up to the size of the earth, and above. (Vol. I, 19–2)

Hence it is that classical mechanics, *qua* special science, provides the basis for most scientific and engineering-based investigations into and treatments of ordinary objects.

Nor does the transition from force-based to energy-based mechanics—in particular, as was advisable for purposes of characterizing quantum phenomena—pose any in-principle difficulty for taking classical mechanics to be a special science in good standing. In particular, as I argue in detail in Wilson 2007, forces and energies are interdefinable, and even if fundamental quantum processes involve (e.g.) particle exchanges rather than Newtonian forces (which is not entirely clear,

insofar as analogical descriptions of such exchanges tend to involve forces), it remains that Newtonian forces can be seen as nonfundamental goings-on which, along with ordinary objects, are properly seen as part of the subject matter of classical mechanics.[5]

To be sure, there remain further questions about the ontological status of Newtonian forces, which have periodically been the target of anti-realist concerns. In Wilson 2007, I address and respond to these concerns; to enter into these details at this point would take us too far afield, however, so I invite the interested reader to attend to that other discussion. Here I will follow Feynman and many other scientists in seeing classical mechanics as a special science in good standing, associated with distinctive laws of nature. Seeing classical mechanics in this light provides our first route to the Weak emergence of some actual ordinary objects—namely, classical ordinary objects—by lights of a DOF-based account.

The classical limit

Classical objects are associated with the so-called 'classical limit', in which (as in Feynman's remarks above) certain quantum features cease to be relevant to the properties and behaviors of macro-objects, in ways that are plausibly interpreted as involving an elimination (and not just a reduction or restriction) of quantum DOF.

For example, consider DOF associated with quantum spin. Classical objects are ultimately composed of quantum entities, whatever exactly these entities may be; but the characteristic states of classical objects do not functionally depend on the spins of the quantum components of these entities. Hence notwithstanding that the values of quantum parameters may in some cases lead to macroscopic differences—for example, readings on a measurement apparatus, and the like, as in the case of Schrödinger's cat—it remains the case that DOF such as quantum spin are eliminated (and not just reduced or restricted) from those needed to characterize entities of the sort appropriately treated by classical mechanics. Hence for reasons discussed in Ch. 5 (§5.2.4), classical objects satisfy the *Proper Subset of Powers Condition*, and more generally are appropriately seen as satisfying the conditions in a DOF-based account of Weak emergence.

Another potential and very general source of elimination in quantum DOF in the classical limit reflects that the probabilistic values of quantum mechanical observables average out to their mean values in this limit, as per the main strategy for understanding how classical mechanics 'emerges' from quantum mechanics

[5] At least this is so for non-gravitational Newtonian forces. The prospects for seeing gravitational forces as real, nonfundamental goings-on are less promising, at least given a General Relativistic understanding of gravitational influence in terms of inertial motion.

(see, e.g., Messiah 1970, 215). Forster and Kryukov (2003) provide a useful explanation by analogy of how this occurs:

> It may be surprising that deterministic laws can be deduced from a probabilistic theory such as quantum mechanics. Here, curve-fitting examples provide a useful analogy. Suppose that one is interested in predicting the value of some variable y, which is a deterministic function of x, represented by some curve in the $x-y$ plane. The problem is that the observed values of y fluctuate randomly above and below the curve according to a Gaussian (bell-shaped) distribution. Then for any fixed value of x, the value of y on the curve is well estimated by the mean value of the observed y values, and in the large sample limit, the curve emerges out of the noise by plotting the mean values of y as a function of x.
>
> To apply the analogy, consider x and y to be position and momentum, respectively, and the deterministic relation between them to be Newton's laws of motion. Then it may be surprising to learn that Newton's laws of motion emerge from quantum mechanics as relations between the mean values of quantum mechanical position and quantum mechanical momentum. These deterministic relations are known as Ehrenfest's equations. In contrast to curve fitting, the Heisenberg uncertainty relations tell us that the quantum mechanical variances of position and momentum are not controllable and reducible without limit. Nevertheless, it is possible for both variances to become negligibly small relative to the background noise. This is the standard textbook account of how Newton's laws of motion emerge from quantum mechanics in the macroscopic limit. (1040)

It is natural to read references to probabilistic variances becoming 'negligibly small' as indicating that the DOF encoding such variances are eliminated in the classical limit, and so as providing support for classical objects' being Weakly emergent—at least, on the assumption that the probabilistic aspects of quantum theory correspond to objective features of quantum goings-on. This is true for many interpretations of quantum mechanics, including versions of a Copenhagen (or 'orthodox') interpretation on which quantum probabilities are grounded in objective properties of particles to be, e.g., in a certain location if measured, and versions of a spontaneous collapse theory, à la Ghirardi *et al.* 1986, on which quantum probabilities are DOF of the wave function, but where these DOF are understood as tracking properties of the associated quantum entities.[6] It may also be possible to see probabilistic quantum DOF as eliminated even if the

[6] Hence Frigg (2009) notes, "A crucial assumption of the theory is that hits occur at the level of the micro-constituents of a system [e.g.,] at the level of the atoms that make up the marble" (267). See also Monton's (2004) 'object' interpretation of a spontaneous collapse theory.

probabilities are 'epistemic', if what this comes to is something like statistically random behavior, and the properties and/or behavior of (some) quantum entities is functionally dependent on such randomness.

To sum up: various ordinary objects of the sort appropriately treated by classical mechanics have DOF required to specify the objects as being in a given characteristic state relevant to the object's law-governed properties (including behavior) that are eliminated as compared to the systems of lower-level entities upon which they cotemporally materially depend, and hence are at least Weakly emergent by lights of a DOF-based account.[7] Now, as above, for an object to be Weakly emergent it is required not just that at least one of its features satisfy the conditions in the schema for Weak emergence, but that the remaining features and behaviors of the object are such as to either also satisfy the conditions in the schema, or else be appropriately identified with (i.e., reducible to) lower-level goings-on. On the assumption that a given ordinary object is in fact appropriately treated by classical mechanics, this further constraint is plausibly in place, since otherwise the object would have new powers to affect lower-level physical goings-on in ways not entirely encoded (just) in classical mechanics. Hence we may say, conditionally, that any object that is appropriately treated by classical mechanics is Weakly emergent, *simpliciter*. Insofar as many special-science objects, including rocks, planets, and the like, are reasonably supposed to be appropriately treated by classical mechanics, all such entities are reasonably supposed to be actually Weakly emergent, by lights of a DOF-based account.

6.1.2 Sortal features and functional realization

Artifacts are commonly taken to be associated with functional roles that to some significant extent serve to characterize what it is to be an object of the type in question. Hence Mitchell (2012) says:

> A chair is describable as an artifact designed to function as something suitable for humans to sit on. It is also the case that chairs are always made out of something material: wood, iron, plastic, etc. So a particular chair could also be described in terms of its material components. Chairs in general are described by their capacity to function as something humans use for sitting. Chair-function is realized by an individual entity's physical components and structure. (174)

[7] As discussed in Ch. 5 (§5.2.4), the eliminations in DOF at issue pertain to the system of composing entities e_i even when the latter are constrained to stand in the relations relevant to their composing a given higher-level entity, since the lower-level laws applying to the e_i, whether constrained or unconstrained, require the full complement of lower-level DOF.

And Searle (2007) says:

> Many of the most common concepts that we use in dealing with the world, for example, concepts like 'cars' and 'bathtubs' and 'tables' and 'chairs' and 'houses', involve the assignment of function. (8)

(While Searle speaks here of concepts, he also clearly takes the notion of function as applying to the referents of the concepts.)

The view that artifacts are functionally characterized no doubt reflects that artifacts are typically created with some purpose in mind. Here I want to develop, in addition, another route to seeing artifacts as functionally characterized, that proceeds by attention to what are sometimes called 'sortal features' (or 'primary kinds', as in Baker 1993) of such objects. Lowe (2007) observes that "The idea that objects are sortally individuated has a long and distinguished pedigree" (514); see also, e.g., Wiggins 2001, 150–1 and Lowe 1989, 11–13. Candidate sortal features for ordinary objects of the varieties at issue here would be features expressing membership in the category at issue, such as 'being a table' or 'being a statue'.

Why is attention to sortal features useful for purposes of assessing whether a given object is metaphysically emergent? To start, note that while there are different understandings of sortal features (see Grandy 2016 for discussion of several definitions and applications), the common thread is that a sortal feature encodes various conditions characterizing the object, at both type and token levels. At the level of types, an object's sortal feature specifies its 'individuation conditions', encoding what distinguishes objects of the type in question from other objects or types of entity. At the level of tokens, an object's sortal feature specifies its 'identity conditions', encoding what it takes for a particular object of the type to be identical to another particular object of the type, either at a time or over time. Identity conditions applying at a time are useful for purposes of counting how many objects of the type associated with the sortal are in a given collection (how many chairs are in the room right now?). Identity conditions applying over time—sometimes called 're-identity' or 'persistence' conditions—are useful for purposes of determining whether an object of a given type has persisted through time, and more importantly, change (is this the same chair as was in the room yesterday?).

It is the persistence conditions encoded in a given sortal feature that I want to highlight as relevant for present purposes. To start, on a standard understanding of the persistence conditions of artifacts, artifacts are compositionally flexible, in that objects such as tables and statues can survive some (and sometimes quite considerable) changes to the lower-level configurations (pluralities or structured aggregates) upon which they depend. Relatedly, it is frequently supposed that artifacts have compositionally flexible material origins, such that, e.g., a given table ('Woody') might have originated from a block of wood somewhat different from

that from which it actually originated (see, e.g., Kripke 1972/80 and Salmon 1989); such 'modal' compositional flexibility might also be encoded in the sortal features of artifacts, as reflecting the conditions under which such objects[8] are identical across worlds, as well as times.

Now, given that the sortal features of artifacts typically encode that such objects are compositionally flexible, both temporally and modally, the question naturally arises: why so? What explains why artifacts are typically compositionally flexible? And here a plausible answer, supporting the sort of characterizations of artifacts mentioned above, is that artifacts are functionally characterized, in that what it is to be an object of the sortal type at issue is to be an object that plays or is capable of playing a specific functional role. For if artifacts are associated with functional roles, and these functional roles are sufficiently abstract that they can be implemented by multiple lower-level configurations, then the compositional flexibility of artifacts would follow as a matter of course—effectively, as an objectual variation on the theme of multiple realizability as usually applied to features. So both the intuitive descriptions of artifacts, and a natural metaphysical basis for the compositionally flexible persistence conditions of artifacts, support taking artifacts to be functionally characterized (again, such that what it is to be an object of the sortal type at issue is to be an object that plays a specified functional role); and I will henceforth assume that this account of artifacts is correct.

Now, the fact that a given artifact is properly seen as functionally characterized does not in itself guarantee that the artifact is Weakly as opposed to Strongly emergent, for reasons having to do with the bearing of mentality on the intentional specification and social implementation of the functions at issue. I'll expand on this further issue down the line. For now, it remains that such a characterization pushes towards thinking that the object is at least Weakly emergent.

To see this, first note that while numbers or other abstracta may be functionally characterized in ways that cannot (or at least cannot obviously) be cashed in causal terms, the functional roles at issue in the case of artifacts are more specifically causal roles. Hence Rosenberg (1994) says:

> Causal descriptions are often called 'functional' role descriptions in the philosophy of science, and I will use this terminology hereafter, understanding 'functional' to mean simply "role in a network of causes and effects". Thus, the functional description of an [axe] identifies it by its functional role, the causes and effects into which it typically enters. An [axe] is a tool [...] for cutting wood (that is, its effect when applied with sufficient momentum will be to sever wood into smaller pieces). (23–4)

[8] Or their counterparts, if an object cannot exist at more than one world, as per Lewis 1986.

Given that the functional roles associated with artifacts (e.g., tables) are more specifically causal roles, it follows that what it is to be an artifact of a certain sortal type is to be an object characterized by a certain causal role, and in particular, to be an object with certain powers. These powers are encoded in the sortal feature of the artifact which, as Lowe put it, individuates the artifact. Hence at this point we can return to a focus on the features of a given object as the potential locus of the metaphysical emergence of the object itself.

Now, independent of the possibility that some of the powers associated with the sortal feature of a given artifact are fundamentally novel in a way indicative of Strong emergence, it seems likely that many of the powers of a sortal feature will be token-identical with the powers of whatever lower-level configuration (plurality or structural aggregate) serves as the dependence base for the artifact on a given occasion. Indeed, it is commonly assumed that all of the powers of artifacts are ultimately powers of their dependence base configurations. Hence Rosenberg (1994) continues:

> Now consider the class of objects that meet the purely functional characterization of an [axe]. They all have to be material objects and have some structural composition or other. But nothing in the functional definition requires that their handles be made of wood or that their heads be made of steel. [...] An [axe] is thus "nothing but" the material out of which it is composed, even though two different axes may share no fact of structural composition in common (above the atomic level). (24)

Granting that an axe is "nothing but" its material base on a given occasion, every power of the axe (or associated sortal feature) must be identical with a power of that base (or associated sortal feature) on that occasion. Merricks (2003) is even more explicit, saying "a baseball and its constituent atoms cannot do any more than those atoms all by themselves" (59), and later, "everything that a baseball causes is caused by its parts at some level of composition" (62). Even some of those who explicitly take into account the role of human intentions in the creation of or social practices involving artifacts seem to assume that an artifact's lower-level dependence base is "capable of fulfilling the functions" of artifacts, as in these remarks of Thomasson (2007):

> [I]f you think that the rules and practices that make up the game of baseball, and intentions of those who rearrange atoms into appropriate spherical shapes, are necessary for there to be baseballs, consider that for atoms to be arranged baseballwise requires atoms tightly bonded in a spherical shape of such-and-such diameter, such that they are jointly capable of fulfilling the functions of baseballs, [i.e., being such that they are] bonded by people with intentions to make baseballs

that meet major league regulations, are usable and to-be-used in such games, and so on. (17)

Supposing that the powers of (the sortal feature) of a given artifact do not go beyond those had by its dependence base configuration, as the physicalist assumes, we are a short step from seeing the artifact as Weakly emergent. For as above, and as per the compositional flexibility of artifacts, these powers are associated with a comparatively abstract functional or causal role, which can be implemented by diverse base-level goings-on. Consider a baseball, for example. It cannot be among the powers of a given baseball (or its sortal feature) to produce a certain precise weight reading on a scale, since given its compositional flexibility, the baseball might exist in the same extrinsic circumstances, and yet, if its dependence base configuration is different, not produce the same reading; by way of contrast, the dependence base configuration is not so compositionally flexible, and so it (or its sortal feature) does have that power. It follows that the baseball (or its sortal feature) has, on a given occasion, *fewer* powers than its dependence base configuration (its sortal feature) on that occasion, and so (given the empirically supported satisfaction of the dependence condition) satisfies the conditions in *Weak Emergence*. And the same is true, more generally, for any compositionally flexible artifact. Since artifacts are typically (always?) compositionally flexible, artifacts are typically (always?) at least Weakly emergent.

It may be that the previous considerations also support, mutatis mutandis, taking natural ordinary objects to be at least Weakly emergent. After all, special-science objects, ranging from cells to mountains, tectonic plates, and planets, are also commonly functionally characterized, and fall under sortal types encoding their compositional flexibility (cf. Lowe's 2007, 519, remarks that "sortal persistence conditions [are] conditions determining what changes an object of any given sort may or may not undergo, as a matter of natural law, while remaining an object of that same sort"). Indeed, according to Rosenberg (1994), "[t]he complexity of nature above the level of the molecule is the result of selection for function and its blindness to structure" (55). It is also worth observing that in the case of natural objects, the functional specification occurs as a result of natural processes or laws as opposed to intentions or social practices, sidestepping the primary basis for thinking that functionally characterized objects might be Strongly emergent.

Be all this as it may, since the primary examples of functionally specified special-science objects are drawn from the biological sciences, and we are here focused on nonliving ordinary objects, I won't here develop this line of thought, but will rather move on to a different reason for thinking that both artifacts and natural ordinary objects are at least Weakly emergent, which has not yet been discussed in the literature.

6.1.3 Metaphysically indeterminate boundaries

Ordinary objects, of either natural or artifactual varieties, typically appear to have indeterminate—that is, imprecise—boundaries. As Tye (1990) observes, "common sense has it that the world contains countries, mountains, deserts, and islands [...] and these items certainly do not appear to be perfectly precise" (396). And as has been frequently noted, even the seeming distinctness of spatial boundaries of ordinary objects such as rocks, tables, and statues, dissolves upon closer examination into an array of lower-level constituents (e.g., atoms) of decreasing but non-zero concentration in the region of a given macro-object boundary—a phenomenon which again suggests that such boundaries are in some sense indeterminate.

How should such seeming boundary indeterminacy be understood? There are three main strategies. The first takes indeterminacy to have its source in how we represent the world (this is 'semantic' indeterminacy); the second takes indeterminacy to reflect the limits of our knowledge of the world (this is 'epistemic' indeterminacy); the third takes indeterminacy to have its source, somehow or other, in the world itself (this is 'metaphysical' indeterminacy). While some sorts of indeterminacy may be best treated in (merely) semantic or epistemic terms, object boundary indeterminacy is most naturally treated in metaphysical terms. It is unclear how indeterminacy in ordinary object boundaries might be a semantic matter, reflecting that we haven't gotten around to drawing certain precise lines: we are not inclined to draw such lines—most importantly, because attribution of any particular precise boundary to an ordinary object (e.g., a mountain, or a table) would be inappropriately arbitrary. Such arbitrariness renders it similarly implausible to take seeming boundary indeterminacy to reflect our inability to discern which perfectly precise boundary is in fact possessed by the object at issue, even granting (what might be questioned[9]) that the boundaries of the lower-level configurations upon which ordinary objects depend are perfectly precise.

Though boundary indeterminacy is most naturally treated in metaphysical terms, many have found metaphysical indeterminacy problematic, as incoherent (Evans 1983), "not properly intelligible" (Dummett 1975), or underspecified (Lewis 1986). Lately, however, two specific accounts of metaphysical indeterminacy have been advanced, each of which is claimed to be coherent and intelligible. In what follows, I'll first argue that one of these—the account initially presented

[9] Various considerations call into question the common assumption that fundamental physical entities, either individuals or micro-configurations, have perfectly precise boundaries. Perhaps the most salient consideration concerns the quantum value indeterminacy associated with position and momentum as incompatible observables, which pushes against taking lower-level entities or configurations to have precise locations, on pain of an infinite variance in momentum. See Calosi and Wilson 2018 for discussion.

and defended in Wilson 2013*a*—is better than the other, so far as treating the indeterminacy of ordinary object boundaries is concerned. I'll then show that the specific preferred treatment entails that ordinary objects are at least Weakly emergent, by lights of a determinable-based account of Weak emergence.

'Meta-level' and 'object-level' accounts of metaphysical indeterminacy

Here I very briefly present the two main approaches to metaphysical indeterminacy, highlighting certain of their differences along the way; see Wilson 2013*a* and 2016*a* and Calosi and Wilson 2019 and forthcoming for more detailed comparisons.

At a general level, the two approaches differ structurally as regards where worldly indeterminacy is supposed to be located. On the metaphysical supervaluationist approach endorsed in Akiba 2004, Barnes 2006, Barnes and Williams 2011, and elsewhere, what it is for a given state of affairs—e.g., an object's having a property—to be metaphysically indeterminate is for it to be metaphysically indeterminate which precise state of affairs, out of some range of candidate precise states of affairs, is actually the case. Here the operating assumption is that all states of affairs are perfectly precise, and worldly indeterminacy consists in the world's being unsettled about which precise state of affairs is in fact the case; indeterminacy is thus structurally located at the 'meta-level', so to speak, as indeterminacy among precise options. On the contrasting approach to metaphysical indeterminacy presented and endorsed in Wilson 2013*a* (see also Bokulich 2014, Wilson 2016*a*, and Calosi and Wilson 2019 and forthcoming), what it is for a given state of affairs to be metaphysically indeterminate is for the constitutive object (more generally, entity) of the state of affairs to have an irreducibly indeterminate property (more generally, feature), where this is understood, in turn, in terms of the constitutive object's having a determinable property, but no *unique* determinate of that property. Such a failure of unique determination can happen in two ways: first, if many candidate determinates of the determinable are instantiated, in such a way that no one determinate is non-arbitrarily taken to be the 'unique' determinate of the determinable instance (this is 'glutty' indeterminacy); second, if no determinate of the determinable is instantiated (this is 'gappy' indeterminacy). Here the operating assumption is that some states of affairs are irreducibly imprecise, and worldly indeterminacy consists in a certain pattern of determinables and determinates (namely, an object's having a determinable but no unique determinate of that determinable); hence indeterminacy is structurally located at the 'object-level', so to speak, in states of affairs themselves.

In a bit more detail: on a meta-level supervaluationist approach, what it is for the world to be 'unsettled' between various precise options is a primitive matter; proponents aim, however, to render primitive metaphysical indeterminacy intelligible by offering a logic and semantics of such indeterminacy along lines

of the logic and semantics offered on a supervaluationist approach to semantic indeterminacy. On a semantic supervaluationist approach, indeterminacy reflects our language's failing to determine a precise extension for certain expressions (e.g., 'bald'), rendering the truth of certain claims (e.g., 'Bob is bald') indeterminate; this failure is then modeled as entailing that precisifications of our language, each compatible with existing (determinate) use, disagree about the extension of the relevant term(s), such that, e.g., according to one linguistic precisification 'Bob is bald' is true, whereas according to another linguistic precisification 'Bob is bald' is false. On the metaphysical version of a supervaluationist approach, it is the world rather than language that is unsettled about whether a given state of affairs obtains, rendering the truth of certain claims (e.g., 'Mount Everest has *B*', for some precise boundary *B*) indeterminate; this failure is then modeled as entailing that 'precisificationally possible worlds', each compatible with existing (determinate) facts, disagree on whether the state of affairs obtains, such that, e.g., according to one precisificationally possible world 'Mount Everest has *B*' is true, whereas according to another precisificationally possible world, 'Mount Everest has *B*' is false. Note that on both semantic and metaphysical supervaluationist accounts, meta-level indeterminacy gives rise to sentential or propositional indeterminacy, and so to the need to introduce an indeterminacy operator into the semantics.

As for an object-level determinable-based approach to metaphysical indeterminacy, the account I endorse (Wilson 2013*a*) is more specifically as follows:

> *Determinable-based MI*: What it is for a state of affairs to be metaphysically indeterminate in a given respect *R* at a time *t* is for the state of affairs to constitutively involve an object (more generally, entity) *O* such that (i) *O* has a determinable feature *P* at *t*, and (ii) for some level *L* of determination of *P*, *O* does not have a unique level-*L* determinate of *P* at *t*. (366)

One might wonder: how can it be that a determinable property is not uniquely determined? It has standardly been assumed (e.g., by Funkhouser 2006, among many others) that it is a requisite feature of determinables that, for every level of specification *L* at which they may be determined, a determinable instance at a time has one and only one determinate at that time. However, as I argue in Wilson 2013*a* and 2016*a* (see also Wilson 2017 for historical discussion), the standard supposition that determinables must be uniquely determined reflects an overly quick generalization from certain paradigm cases of determinable and determinate instances, and should be rejected as generally characterizing determinables and determinates.

Indeed, there is a case to be made that even instances of color, the feature most often highlighted as illustrative of a determinable that can be further determined, are not always uniquely determined. In particular, the case of an iridescent feather,

the color of which is relative to perspective (the feather is red from one perspective, blue from another), is reasonably interpreted as one where the feather is colored (has the determinable *color*) but is not any unique (one and only one) color (does not have a unique determinate of *color*), primarily because any attribution of a specific shade (e.g., red rather than blue) as its purported unique color would be inappropriately arbitrary. At best, one can say that the feather is red relative to one perspective (more generally, circumstance), and is blue relative to another. This and other cases of what I call 'multiple relativized determination' provide good reason for thinking that the conditions in *Determinable-based MI* can be satisfied—in the case of a feather, in 'glutty' fashion (again, adverting to there being too many candidate determinate instances on the scene).[10]

There is, of course, much more to be said about each approach, and I will say a bit more down the line; but this sketch will be enough to proceed.

Application to the case of ordinary object boundaries

I now want to apply the two accounts to a candidate case of indeterminacy in an ordinary object boundary, and argue that a determinable-based account better treats the case.

On a supervaluationist account, such indeterminacy is treated as follows:

> *Supervaluationist MI* (ordinary object boundaries): What it is for an ordinary object *O* to have an indeterminate boundary at a time *t* is for it to be primitively metaphysically indeterminate at *t* which precise boundary *O* has (i.e., for it to be primitively metaphysically indeterminate at *t* which precisificationally possible world is actual).

Here, what it is for Mount Everest to have an indeterminate boundary at a given time is for the world to be unsettled at that time about which precise boundary Mount Everest has—or, in terms of precisificationally possible worlds, for it to be primitively indeterminate which such possible world is actual. Notwithstanding that it is unsettled at *t* which precise boundary Mount Everest has, it is nonetheless settled (since 'super-true': true at every precisificationally possible world) that Mount Everest has a precise boundary, and it is moreover settled (for the same reason) that for every precise boundary *B*, either Mount Everest has *B* or Mount Everest does not have *B*. And similarly for the indeterminate boundaries of other natural and artifactual ordinary objects.

On a determinable-based account, such indeterminacy is treated as follows:

[10] Open future cases (see Wilson 2016*a*) and some interpretations of certain quantum phenomena (see Bokulich 2014, Wolff 2015, and Calosi and Wilson 2019) provide motivation for taking the conditions in *Determinable-based MI* to be satisfiable in 'gappy' fashion, whereby no determinates are available to provide the unique determinate of the determinable.

> *Determinable-based MI* (ordinary object boundaries): What it is for an ordinary object *O* to have an indeterminate boundary at a time *t* is for *O* (i) to have a determinable boundary property *P* but (ii) for some level *L* of determination of *P*, not to have a unique level-*L* determinate of *P* at *t*.

For example, on a determinable-based account, what it is for Mount Everest to have an indeterminate boundary at a given time is for Mount Everest to have a determinable boundary property at that time, but not to have a unique determinate of that boundary property at that time. What it is for a table (call it 'Woody') to have an indeterminate boundary at a time is for Woody to have a determinable boundary property at that time, but no unique determinate of that property at that time. And similarly for other natural and artifactual ordinary objects.

Why think that mountains, tables, and other ordinary objects can have a determinable boundary property at a time, but no unique determinate boundary property at that time? And given that such ordinary objects satisfy the conditions in *Determinable-based MI*, is the indeterminacy at issue glutty or gappy? On the view I endorse, the indeterminacy of ordinary object boundaries (at a time; henceforth I'll leave this qualification implicit) is plausibly seen as reflecting that there are multiple candidate determinate boundary properties, associated with different and typically overlapping micro-configurations (pluralities or structural aggregates) in the vicinity of the ordinary object's boundary, such that it would be inappropriately arbitrary to pick out just one of these determinate boundary properties as being the unique determinate such boundary had by the ordinary object. This sort of case, involving multiple relativized determination, involves glutty, not gappy, indeterminacy.

Consider Mount Everest. This mountain has the determinable property of having a boundary *around here* (gesturing at the general vicinity of the mountain). Now, does it make sense to take Mount Everest to have some unique maximal determinate of this determinable boundary property? It's hard to see how. If Mount Everest were to have a unique maximally determinate boundary property, it would presumably be inherited, either by way of identity or realization, from the (presumed) maximally determinate boundary property of some micro-configuration in its vicinity. But at any given time, there are multiple micro-configurations in the vicinity of Mount Everest. These overlap as regards the clear interior regions of Mount Everest, but differ in their spatial extent: some have more dirt, some less, some are bigger, some smaller. Which micro-configuration is 'the' micro-configuration whose (presumed maximally determinate) boundary is possessed by Mount Everest?

There is not, it seems, any good answer to this question except one according to which the question is ill-formed. For none of the candidate micro-configurations has any more metaphysical claim to be considered 'the' dependence base of Mount

Everest than any other. Correspondingly, taking Mount Everest to have, as its purported unique maximally determinate boundary, the boundary of any one of these micro-configurations would be inappropriately arbitrary.[11]

Summing up: Mount Everest has a determinable boundary property, but (owing to the presence of too many candidate maximal determinates of this determinable) does not have a unique maximal determinate of this property; hence the state of affairs of Mount Everest's having a boundary satisfies the conditions in *Determinable-based MI*.[12] And similarly for the indeterminate boundaries of other natural and artifactual objects.

Now, in previous work I've highlighted several advantages of a determinable-based approach to metaphysical indeterminacy over a metaphysical supervaluationist approach, including that while a supervaluationist account introduces sentential or propositional indeterminacy, and so requires some revisions to both classical logic and semantics (see Williamson 1994, Ch. 6), a determinable-based account does not introduce sentential or propositional indeterminacy, and so requires no such revisions.[13] Here I want to highlight two important advantages specifically applying to the case at hand.

The first advantage is that a supervaluationist treatment of object boundary indeterminacy is, unlike a determinable-based approach, not clearly intelligible. Recall that on a supervaluationist approach, metaphysical indeterminacy is

[11] This observation is along lines of what Unger (1980) evocatively calls 'the problem of the many', according to which attempts to identify any given macro-object (in his preferred case: Tibbles the cat) with some lower-level configuration runs into the problem that there are many equally good candidates, at a given time, for being the lower-level dependence base entity, such that the identification of the macro-object with any one of these candidates would be inappropriately arbitrary. The problem of the many can also be seen as an interesting variation on cases of multiple realizability, where a given feature that is in fact realized by a given micro-feature might possibly be alternatively realized by a different micro-feature. In the case of Mount Everest, the multiplicity of realization of the determinable boundary property is actual, not merely possible; and as in Unger's problem, any attempt to pick out one of the micro-configurations in the vicinity of Mount Everest and attribute its maximally precise boundary to Mount Everest as the unique determinate of the mountain's determinable boundary property would be inappropriately arbitrary.

[12] Again, the satisfaction of the conditions in *Determinable-based MI* in the cases at hand is indicative of glutty rather than gappy indeterminacy. Like the color of an iridescent feather, the determinable boundary of Mount Everest admits, at a time, of multiple determinations, each associated with one of the micro-configurations in the vicinity. As with the case of the feather, some may see it as appropriate to attribute the precise boundaries to the mountain, albeit in relativized fashion; others may maintain that even relativized attributions of precise boundaries are inappropriate unto characterizing ordinary objects. This is a choice point; but either way, in cases of such multiple relativized determination, it is not an option to attribute a unique, or indeed any, precise boundary to the mountain (more generally, ordinary object) in unrelativized fashion.

[13] Consider our test case, involving Mount Everest's indeterminate boundary. On this treatment, it will be true that Mount Everest has a boundary, and false that Mount Everest has a perfectly precise boundary. It will be false, for every unrelativized attribution of such a precise boundary, that Mount Everest has that boundary. It may or may not be true, depending on the choice point mentioned in note 12 and on the specific facts at hand, that Mount Everest has such a precise boundary in relativized fashion. And so on. Similarly, all the usual theorems of classical logic and semantics hold.

primitive, but is supposed to be rendered intelligible by analogy to semantic supervaluationism. But at least for the case at hand, a metaphysical supervaluationist account does not naturally inherit the intelligibility of its semantic precursor; for though it is clear enough what it would be for our language to be unsettled between various ways of linguistically precisifying a vague term such as 'bald', it is less clear what it would be for the world itself to be unsettled, at a given time t, between various precise ways it might be at t—especially given that (as we have seen) the logic of precisificational possibilities entails that it *is* settled at t that the world is one of those precisificationally precise ways. As applied to the case of Mount Everest: how could it be that the world is settled, at t, that Mount Everest has a precise boundary, but is unsettled, at t, about which boundary that is? Conversely, if the world is unsettled at t about which precise boundary Mount Everest has, how could it be settled at t that it has a precise boundary?[14] By way of contrast, a determinable-based account of object boundary indeterminacy is clearly intelligible; for on this view indeterminacy is not primitive, but is rather *reducible* to a certain pattern of features of the sort with which we are already experientially and theoretically familiar, and which pattern is clearly intelligibly modeled (by, e.g., such mundane cases as an iridescent feather).[15]

The second advantage is that a supervaluationist account incorrectly presupposes that it would be appropriate, in principle, to assign a perfectly precise boundary to an ordinary object such as Mount Everest, in unrelativized fashion. Relatedly, on a metaphysical supervaluationist account, a claim such as 'Mount Everest has a unique, maximally determinate boundary' comes out true (since 'super-true': true in every precisificationally possible world). A metaphysical supervaluationist approach thus fails, like semantic and epistemic approaches to boundary indeterminacy, to accommodate intuitions according to which it is part of what it is to be Mount Everest, or pretty much any other ordinary object, to *not* have a precise boundary. By way of contrast, a determinable-based account accommodates this intuitive central feature of ordinary objects.

I take it that, even setting aside the general advantages of a determinable-based account, these considerations provide good reason to prefer a determinable-based treatment of ordinary object boundary indeterminacy.

[14] Here it is worth comparing a better case for the supervaluationist—the case of the open future. I can see how the world might be settled at t that, at some future time t', either a sea battle will be going on or a sea battle won't be going on, but unsettled at t about which of these options will be the case at t'. The difficulty attaching to the case of ordinary object boundary indeterminacy lies in understanding how to apply this pattern of worldly settledness and unsettledness to a case of present-tense indeterminacy.

[15] In reductively analyzing metaphysical indeterminacy using off-the-shelf resources, a determinable-based account is also more parsimonious than a metaphysical supervaluationist account; but since (as per Ch. 1, §1.4.5) I don't place much weight on parsimony considerations, I won't press this advantage.

The connection to Weak emergence

The previous result provides a basis for taking the vast store of ordinary objects, which typically have indeterminate boundaries, to be at least Weakly emergent, that proceeds by looking more closely at the analogy between multiple determination and multiple realization, and considering its implications. As above, on a determinable-based treatment, the metaphysical indeterminacy in ordinary object boundaries is of the glutty variety: the determinable feature is associated with multiple candidate determinates of the determinable at a given time, each associated with a micro-configuration in the vicinity, such that the determinate boundary properties can be had by the ordinary object, at best, in relativized fashion. This phenomenon of multiple relativized determination echoes the phenomenon of multiple realizability, with the main difference being that in cases of multiple determination the multiplicity is actual, rather than merely possible.[16] Now, as previously discussed, while there are reductive strategies for accommodating cases of multiple realizability, attention to the relations between the powers associated with a multiply realized feature and the powers associated with any one of its realizing features suggests that reductive strategies fail, since multiply realized types have fewer powers than each of their realizing types (reflecting effects that the latter can cause that depend on differences between the realizing types), and this proper subset relation between the powers of the realized and realizing types will be inherited by instances of the types, satisfying the *Proper Subset of Powers Condition* and blocking reduction at both the type and token levels.

Here I want to observe that in cases where a given determinable admits of multiple (relativized) determination, the determinable type will also be reasonably taken to have fewer powers than each of the more determinate types, again reflecting effects that the latter types can cause that depend on differences between determinate types, at a given level of specification. For already-rehearsed reasons, this proper subset relation as holding between powers of determinable and determinate types will also hold between tokens of those types, hence satisfying the *Proper Subset of Powers Condition*. That the multiple determinates are all actually instantiated (albeit in relativized fashion) in the case of the determinable boundary properties of ordinary objects, as opposed to being merely possibly instantiated (in non-relativized fashion) in typical cases of multiple realization, doesn't affect the relation between powers of the higher-level feature and lower-level feature at issue. Hence given that the determinable boundary properties of ordinary objects satisfy the conditions in *Determinable-based MI* in glutty fashion (as involving multiply relativized determination), and coupled with associated satisfaction of the cotemporal material dependence condition, it will follow that the determinable

[16] Even this difference is superficial, in that it has been observed that certain features might be actually multiply realized at a time.

boundary property of an ordinary object satisfies the conditions in a determinable-based implementation of *Weak Emergence*, such that the ordinary object at issue will be at least Weakly emergent.

Adding further support to this result is that the connection between realization and the determinable/determinate relation has in past decades been recognized and appealed to (by Macdonald and Macdonald 1986, Yablo 1992, Wilson 1999 and 2009, Shoemaker 2000/2001, and others) as providing the basis for an account of realization (in particular, one allowing that determinable and associated determinate instances may be possessed by different objects).[17] Considerations of how best to treat the metaphysical indeterminacy of ordinary object boundaries thus provide independent support for thinking that ordinary objects are at least Weakly emergent, by lights of a determinable-based account of such emergence.

6.2 Are ordinary objects Strongly emergent?

The considerations of the previous section suggest that many ordinary objects, of both natural and artifactual varieties, are at least Weakly emergent. Might any ordinary objects be, moreover, Strongly emergent? While the Strong emergence of some natural ordinary objects remains an empirical possibility, the best case for Strong emergence here attaches to the case of artifacts. In the remainder of this section I'll briefly lay out this case. I will not (in this chapter) assess this case, since as we will see, whether it goes through ultimately hinges on the status as Strongly emergent, or not, of conscious mentality—a topic to be discussed in the next chapter.

6.2.1 Two routes to the Strong emergence of artifacts

Recall that for an object to be Strongly emergent, it must have at least one feature associated with a fundamentally novel power—a power not token-identical to any power of its base feature on a given occasion. Yet as above, philosophers commonly assume that every power of (every feature of) an artifact, on a given occasion, is uncontroversially a power of (some feature of) the physically acceptable micro-configuration upon which it depends, on that occasion.

[17] Most objections to determinable-based accounts of realization have focused on whether such realization would be suited for accommodating mental-physical realization. I return to this issue in Ch. 7 (§7.2.2).

One might think that this is a reasonable assumption. After all, while it might remain at least an open question whether mental features and their bearers (e.g., persons) have fundamentally novel powers, associated (as on the British Emergentist view, as well as my own interaction-relative version of Strong emergence) with nonphysical fundamental interactions, it's considerably less plausible that any such fundamental interactions enter into determining the existence or behaviors of baseballs, tables, and the like.

This is too quick, however. For artifacts are associated with functional roles in which mentality is deeply implicated, in ways that, were the associated mental features to be Strongly emergent, might open the door to artifacts' having fundamentally novel powers.

Mentality potentially enters into the functional roles of artifacts in two ways: first, mentality enters into the association of a given functional role with a given artifact; second, mentality is, one way or another, typically to some extent constitutive of the role in question. Hence to be a baseball is to be an object that enters into a highly complex set of human practices, ranging from the playing of the game itself and all that mentally entails, to the emotional and even ethical appreciation that is a concomitant of the game, its players, and its values. To be a statue is to be an object that enters into a highly complex set of human practices, ranging from the creation and exhibition of said object, to the perceptual and aesthetic, not to mention economic, interactions that are some part of these practices. And so on. Given that mental features are typically implicated in the characterization of artifacts as having specific functional roles as well as being to some extent constitutive of the roles themselves, we can ask the question: supposing that the mental features in question turned out to be Strongly emergent, would it reasonably follow that the associated artifacts were also Strongly emergent?

In general, it seems to me, the bare involvement of a given mental state in either of these aspects might not be enough to render the associated artifact Strongly emergent. For example, were someone with a Strongly emergent mental state to stipulate that a given rock is to be deemed a 'Faller'—an artifact whose role is to fall when dropped in accord with the laws of nature, their doing so would probably not be seen as sufficient unto bestowing a fundamentally novel power on the Faller. Still, for other functional roles, and contingent on the status as Strongly emergent of normative, intentional, and/or perceptual features, it seems plausible to think that an intentional act of stipulation of a given functional role could reasonably be seen as bestowing upon the artifact in question one or more fundamentally novel powers, as an extension of the idea that the presence of a novel fundamental interaction can serve to ground novel powers. For example, contingent on the status of the associated mental act, perhaps Andy Warhol's stipulation that a given Brillo box was to be considered an artwork suffices to bestow certain Strongly

emergent powers on the box—to produce certain aesthetic responses, to fetch a high price at auction, and so on.

The latter considerations indicate that we can say this much: insofar as various forms of mentality typically enter into both constructing and constituting the functional roles associated with artifacts, and insofar as artifacts are plausibly taken to be to some important extent characterized by the contributions of such mental features, then any features or associated powers of artifacts which are constituted or caused to exist by Strongly emergent mental features would also be appropriately considered Strongly emergent.

It is worth connecting this result to Merricks's conclusions in his (2003). There Merricks argues that while ordinary objects, such as baseballs, do not exist, persons do exist—precisely because, he supposes, persons (in particular, certain of their mental features) have fundamentally novel powers relative to the powers had by their dependence base goings-on. In other words, for Merricks, persons are Strongly emergent. The present discussion shows that there is a potential tension in Merricks's view, at least for the artifacts that are the focus of his discussion; for supposing that the existence and features of artifacts crucially depend on the existence and mental features of persons, the Strong emergence of persons and associated mental features might well be inherited by any artifacts that they create. And in that case, Merricks would be committed to the existence of baseballs and other ordinary objects, after all.

6.3 Concluding remarks

Let's sum up the results of this chapter. I've argued that there are three different cases to be made for the claim that some ordinary objects are Weakly emergent or at least Weakly emergent. First, classical objects—special-science objects of the sort appropriately treated by classical mechanics—are arguably Weakly emergent, by lights of a DOF-based account, thanks to the elimination of quantum DOF in the classical limit. Second, artifacts are arguably at least Weakly emergent, by lights of a functional realization account, thanks to the support for such a treatment that can be extracted from the compositionally flexible persistence conditions typically encoded in the sortal features of artifacts. Third, both natural and artifactual ordinary objects are arguably at least Weakly emergent, by lights of a determinable-based account, thanks to the metaphysical indeterminacy of the boundaries of ordinary objects, which indeterminacy is best treated in determinable-based terms.

Moreover, the possibility remains, especially for artifacts, that these are ultimately not just Weakly but Strongly emergent, owing to the role mentality plays

both in the specification and the constitution of the functional roles which are typically associated with artifacts. As such, the status of any artifacts as Strongly emergent arguably ultimately depends on the status of conscious mentality, to be explored in the next chapter. These considerations more generally suggest that those who are committed to the Strong emergence of persons and certain of their conscious mental features might well find themselves with considerably more entities in their ontology than they are on record as accepting.

Before moving on, I pause to note one other implication of these results. In her (2010), Thomasson discusses a number of metaphysical concerns directed at the supposition that ordinary objects exist, and argues that these concerns arise from inappropriately applying broadly 'scientistic' methodology in service of answering philosophical questions:

> Although the arguments against the existence of ordinary objects do not rely on any particular piece of scientific knowledge, many of them do rely on a certain scientistic approach to metaphysics: the view that metaphysics is of a piece with (and indeed part of the same total theoretical enterprise as) natural science. [...] Lying behind many of the arguments against ordinary objects is the assumption that metaphysics is engaged in explanatory theory construction following the same principles as those governing the natural sciences—or (perhaps better) engaged in one and the same enterprise of constructing the best 'total theory'. (596)

Thomasson takes such an assimilation of scientific methodology to the metaphysical enterprise to be problematic, as giving rise, for example, to concerns about causal overdetermination:

> Straightforward causal redundancy arguments rely on accepting the Eleatic Principle: that we should only accept the existence of those entities that 'make a causal difference'. This may be a reasonable principle for deciding whether or not we should accept the existence of neutrons as well as protons and electrons—theoretical particles posited to explain experimental data. But whether or not it carries over to the question of whether we should 'posit' tables 'as well as' the particles that make them up is less clear. (596)

Thomasson suggests that this sort of concern is effectively a spandrel of a mistaken assimilation of scientific to metaphysical methodology. Correspondingly, she maintains, the correct response to this and other concerns about ordinary objects is to reject this methodological assimilation, and the associated assumption that questions pertaining to the existence and nature of ordinary objects should be

treated in ways consonant with those questions as pertaining to special-science entities.

How to respond? To start, one might think that investigations in the metaphysics of science are not really 'scientistic', at least in that the methodology of metaphysicians of science is not empirical per se. One might also observe that it would be surprising if the metaphysical treatment of special-science entities were required to be different from the metaphysical treatment of ordinary inanimate objects, since after all, many ordinary objects are also (e.g., classical mechanical) special-science entities.

In any case, the results of this chapter show that there is no in-principle reason to think that investigations into the existential (more generally, metaphysical) status of ordinary objects should proceed differently from investigations into the existential (metaphysical) status of special-scientific entities. Thomasson's suggestion to the contrary seems to have been primarily motivated by thinking, first, that the usually stated concerns with ordinary objects (e.g., Kim-style causal overdetermination concerns) arise from trying to give scientific and ordinary objects (including artifacts) a unified treatment, and second, taking the concerns as attaching to scientific goings-on not to admit of any good answers. But as I have argued, there are good responses to the concerns at issue, whether natural or artifactual ordinary objects are at issue. Nothing stands in the way of a systematic treatment of natural and artifactual ordinary objects as at least Weakly emergent, and contingent upon future empirical results and the import of mentality to be next considered, perhaps even Strongly emergent.

7
Consciousness

I turn now to considering whether consciousness or conscious experience of the sort that we and other creatures enjoy is either Weakly or Strongly emergent. There are, to be sure, many forms or species of consciousness, including perceptual awareness of the external world, conscious awareness of internal states (e.g., pain), and self-consciousness—consciousness of ourselves as conscious beings. As Chalmers (1996) notes,

> Conscious experiences range from vivid color sensations to experiences of the faintest background aromas; from hard-edged pains to the elusive experience of thoughts on the tip of one's tongue; from mundane sounds and smells to the encompassing grandeur of musical experience; from the triviality of a nagging itch to the weight of a deep existential angst; from the specificity of the taste of peppermint to the generality of one's experience of selfhood. All these have a distinct experienced quality. All are prominent parts of the inner life of the mind. (4)

Notwithstanding this diversity, little in this chapter hinges on differences between these forms of consciousness; so unless some specific variety is under discussion, I will speak generically of consciousness or conscious awareness (or associated mental features), which may have as its seeming object the external world, one's internal states, or (as a special case of the latter) consciousness itself.[1]

Interestingly, and in some contrast to the cases of complex systems and ordinary objects considered in previous chapters, the cases proferred for the emergence of consciousness are most frequently aimed at showing that consciousness is Strongly rather than merely Weakly emergent. Hence Chalmers (2006*a*) says, "there is exactly one clear case of a strongly emergent phenomenon, and that is the phenomenon of consciousness" (3). In turn, it is often suggested that the main

[1] One respect in which the discussion to follow is not neutral is in supposing that conscious mental states are an apt subject of metaphysical attention, as potentially (in the theoretical sense) reducible to, Weakly emergent from, or Strongly emergent from lower-level physical states. See Hellie 2019*a*, 2019*b*, and forthcoming for development of a competing view on which this supposition is incorrect, and consciousness is rather given an expressivist treatment. Hellie's work challenges certain foundational suppositions of the debates addressed in this chapter, and a full engagement with these challenges must await a different occasion; what follows should accordingly be seen as conditional on there good being responses to these challenges.

Metaphysical Emergence. Jessica M. Wilson, Oxford University Press (2021). © Jessica M. Wilson.
DOI: 10.1093/oso/9780198823742.003.0007

motivation for taking consciousness to be Strongly emergent reflects a commonly acknowledged failure of consciousness to be predictable from or explainable in terms of lower-level physical phenomena. Hence Broad (1925) says:

> [An archangel] would know exactly what the microscopic structure of ammonia must be; but he would be totally unable to predict that a substance with this structure must smell as ammonia does when it gets into the human nose. The utmost that he could predict on this subject would be that certain changes would take place in the mucous membrane, the olfactory nerves and so on. But he could not possibly know that these changes would be accompanied by the appearance of a smell in general or of the peculiar smell of ammonia in particular, unless someone told him so or he had smelled it for himself. (71)

More recently, Bedau (2010) says:

> Our inability to have found any plausible micro-level explanation for conscious mental states might reflect just our ignorance, but another possibility is that these phenomena really are strongly emergent. (60, note 3)

Indeed, existing arguments offered as supporting the Strong emergence of consciousness nearly all rely, one way or another, upon the supposition that consciousness, or certain of its characteristic features or aspects,[2] lies beyond the explanatory reach of any lower-level physical goings-on.

As the reader may suspect, the presence of even an insuperable or 'in-principle' explanatory gap can't be the whole supporting story, however, since—to take one previously discussed example—many complex nonlinear phenomena are clearly physically acceptable, in spite of having features that are, in some relevant sense of 'insuperability' (as, e.g., beyond the access of any empirically possible determination), insuperably unpredictable or otherwise unexplainable in lower-level physical terms. What's moreover at issue in discussions of consciousness is that the explanatory gaps are taken to be metaphysically significant, in reflecting not just broadly mathematical barriers to explanation, such as nonlinearity and associated sensitivity to initial conditions, but rather that certain characteristic features of consciousness—in particular, subjective or qualitative aspects of conscious experience—depart so greatly from lower-level physical features that this divergence provides reasonable grounds for thinking that no physicalist account

[2] Talk of characteristic features (or aspects) of consciousness or conscious states should be understood as broadly neutral on exactly how such features enter into such states. In particular, such talk isn't intended to suggest that characteristic features of conscious states are second-order features of features.

of consciousness of either reductive or nonreductive (Weakly emergent) varieties could possibly be correct.[3]

In being so motivated, explanatory gap arguments in favour of the Strong emergence of consciousness deserve independent consideration. As I will argue, however, these arguments are ultimately uncompelling for reasons not much previously explored, notwithstanding the considerable critical attention that has been given to these arguments. More generally, I will argue that while it remains an open empirical possibility that consciousness is Strongly emergent, at present we have no compelling philosophical or empirical motivation for taking this to actually be so.[4]

On the assumption that consciousness isn't (doesn't turn out to be) Strongly emergent, might it be Weakly emergent rather than ontologically reducible? Here I again present an underexplored reason to endorse a positive answer to this question. The argument proceeds via the claim that qualitative conscious states—e.g., states of conscious awareness of colors or pains—are typically determinable rather than (maximally) determinate, in a way that defensibly renders them suitable (again, assuming that they are not Strongly emergent) for being realized in determinable-based fashion, and hence Weakly emergent.

The overall conclusion is that there are reasons to think that consciousness is at least Weakly emergent from lower-level physical goings-on.

[3] An early expression of this line of thought can be found in Ewing's (1951) reasons for rejecting the identity theory:

> Nineteenth-century materialists were [...] inclined to identify thinking, and mental events generally, with processes in the central nervous system or brain. In order to refute such views I shall suggest your trying an experiment. Heat a piece of iron red-hot, then put your hand on it, and note carefully what you feel. You will have no difficulty in observing that it is quite different from anything which a physiologist could observe, whether he considered your outward behaviour or your brain processes. The throb of pain experienced will not be in the least like [...] anything described in textbooks of physiology as happening in the nervous system or brain. I do not say that it does not happen in the brain, but it is quite distinct from anything that other people could observe if they looked into your brain. [...] We know by experience what feeling pain is like and we know by experience what the physiological reactions to it are, and the two are totally unlike. [...] the difference is as plainly marked and as much an empirical matter as that between a sight and a sound. The physiological and the mental characteristics may conceivably belong to the same substance [...] but at least they are different in qualities, indeed as different in kind as any two sets of qualities. (101–2)

[4] This result applies to the qualitative or subjective conscious features that are the usual focus of discussions of the ontological status of consciousness or conscious experience. As we'll see in the next chapter, I think there is a good existing case for taking libertarian free will to be Strongly emergent; hence if exercises of such free will involve consciousness, then the result here should be accordingly understood as restricted.

7.1 Is consciousness Strongly emergent?

In this section, I consider two different forms of argument that have been or might be offered in support of consciousness's being Strongly emergent: first, 'knowledge arguments' as presented by Nagel (1974) and Jackson 1982 and 1986; second, the 'conceivability' or 'zombie' argument as presented by Chalmers 2009. The qualifier 'might be' reflects that while proponents of these arguments take the conclusions to undermine the physical acceptability of conscious mental states, they do not thereby go on to conclude that consciousness is Strongly emergent.[5] For present purposes, what is important is that the considerations these authors raise can and have been offered in support of taking consciousness to be physically unacceptable, and hence (along with other premises) for taking consciousness to be Strongly emergent, in the sense operative in the associated schema for emergence.

7.1.1 The knowledge arguments

The suggestion that even complete knowledge of physical goings-on might not suffice for knowledge of certain aspects of conscious experience can be found throughout history. Echoes of the basic idea can be found in Leibniz's (1714) 'Mill argument':

> [W]e must confess that perception, and what depends upon it, is inexplicable in terms of mechanical reasons, that is through shapes, size, and motions. If we imagine a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters a mill. Assuming that, when inspecting its interior, we will find only parts that push one another, and we will never find anything to explain a perception. (§17; GP: VI, 609/AG, 215)

Leibniz concluded that perceptual consciousness should be understood as located in a 'simple substance' rather than 'the composite'. As above, the British Emergentist Broad also offered a version of a knowledge argument (Broad 1925, 71), involving a mathematical archangel who knows all the mechanistic facts about chemical goings-on, in support of taking qualitative aspects of conscious

[5] Indeed, Nagel (1979) elsewhere registers his supposition that there is no emergence of a physically unacceptable variety (see van Cleve 1990 for discussion). Both Jackson and Chalmers appear to endorse the possibility of an anti-physicalist form of emergence (albeit of an epiphenomenalist variety) as a further concomitant of their arguments, though Chalmers moreover allows that the anti-physicalist upshot might rather be a form of panpsychism.

experience to be Strongly emergent. More recently, both Nagel (1974) and Jackson (1982 and 1986) have offered arguments aiming to show that one could have complete physical knowledge of some entity or subject matter, but nonetheless fail to know certain facts pertaining to conscious states associated with the entity or subject matter; from this they conclude that physicalism, at least as commonly understood, is false. Such an anti-physicalist conclusion, coupled with the assumption that states of conscious awareness cotemporally materially depend on lower-level physical states, would provide positive motivation for consciousness's being Strongly emergent.

Nagel's discussion proceeds by attention to the question of what we ought to be able to understand about the conscious experience of creatures relatively different from ourselves.[6] He first connects the notion of consciousness to what he calls "the subjective character of experience":

> Conscious experience is a widespread phenomenon. It occurs at many levels of animal life [...] [N]o matter how the form may vary, the fact that an organism has conscious experience *at all* means, basically, that there is something it is like to *be* that organism. There may be further implications about the form of the experience; there may even (though I doubt it) be implications about the behavior of the organism. But fundamentally an organism has conscious mental states if and only if there is something that it is like to *be* that organism—something it is like *for* the organism. We may call this the subjective character of experience.[7]
>
> (436)

Nagel goes on to more specifically consider the case of a bat—a creature plausibly having conscious experience, but of a very different sort from any to which we are privy—and to raise to salience difficulties in our being able to comprehend what it is like to be that sort of creature. As he observes, there is no clear way to extrapolate from our own perceptual experience to that of a creature who navigates the world using echolocation. And nor, crucially, would any amount of knowledge of the physiology of bats, whether pitched at a lower or at a higher level of physical goings-on, enable us to gain knowledge of this comparatively alien form of conscious experience. Nagel concludes that such knowledge is beyond our ken,

[6] See Feigl 1959 for a historical antecedent of attention to this issue, involving a Martian in possession of all the physical facts about humans, who would nonetheless "be lacking completely in the sort of imagery and empathy which depends on familiarity (direct acquaintance) with the kinds of qualia to be imaged or empathized" (431).

[7] For anticipations of such 'what it's like' talk, see also Farrell 1950, 185–8 and Feigl 1967, 139–40. For recent discussion of the locution, see Hellie 2007.

and goes on to suggest that the latter gap spells trouble for any view on which consciousness (and its subjective character) are either ontologically reducible to or realized in (that is, Weakly emergent from) lower-level physical goings-on:

> While an account of the physical basis of mind must explain many things, this appears to be the most difficult. [...] if physicalism is to be defended, the phenomenological features must themselves be given a physical account. But when we examine their subjective character it seems that such a result is impossible. The reason is that every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view. (437)

Nagel sees here a "divergence between [...] two kinds of conception: subjective and objective" (438). An account of consciousness as physically acceptable must include a treatment of the subjective character of consciousness, but on the assumption that physical theory is solely concerned with the objective or third-person point of view, any such treatment will be, he maintains, "impossible".

Jackson's (1986) knowledge argument (for an early variation on the theme, see Jackson 1982) also aims to establish that there is an insuperable, metaphysically significant gap in relevantly comprehensive knowledge of certain physical goings-on and knowledge of what it is like to have certain conscious experiences. Jackson's focus differs from Nagel's in highlighting the gap as attaching to the qualitative aspects of conscious experiences, as per the following thought experiment:

> Mary is confined to a black-and-white room, is educated through black-and-white books and through lectures relayed on black-and-white television. In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in *completed* physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this, including of course functional roles. If physicalism is true, she knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is just what physicalism denies. [...] It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. This is rightly described as *learning*—she will not say "ho, hum." Hence, physicalism is false. This is the knowledge argument against physicalism in one of its manifestations. (291)

As Jackson qualifies, at issue here is not that Mary comes to learn something about her own experiences, but rather that Mary comes to learn something about the experiences of others:

> The trouble for physicalism is that, after Mary sees her first ripe tomato, she will realize how impoverished her conception of the mental life of *others* has been *all along*. She will realize that there was, all the time she was carrying out her laborious investigations into the neurophysiologies of others and into the functional roles of their internal states, something about these people she was quite unaware of. (292, emphases in text)

How might a physicalist best respond to the knowledge arguments? In what follows, I'll focus on Jackson's version; the application to Nagel's argument and other variations on the theme will be clear. I'll moreover focus on the version of Jackson's argument presented in Nida-Rümelin 2015, which helpfully articulates the key premises and conclusions, as follows:

P1: Mary has complete physical knowledge about human color vision before her release.
C1: Therefore, Mary knows all the physical facts about human color vision before her release. (P1)
P2: There is some (kind of) knowledge concerning facts about human color vision that Mary does not have before her release.
C2: Therefore, there are some facts about human color vision that Mary does not know before her release. (P2)
C3: There are nonphysical facts about human color vision. (C1, C2)

The knowledge arguments have generated an enormous amount of literature, and a full treatment of all the variations on, ramifications of, and responses to these arguments is beyond the scope of this chapter (see Nida-Rümelin 2015 for extensive discussion and references). Here I'll present my preferred strategy, which is apparently not much on the books—namely, to deny P1. Along the way I'll positively contrast this approach with certain other more popular strategies.

The 'Incomplete Physical Knowledge' strategy

According to P1 of the knowledge argument, Mary has complete physical knowledge about human color vision before her release. The strategy of denying P1 is not much considered; hence Nida-Rümelin passes it over, saying, "it seems hard to deny that it is in-principle possible to have complete physical knowledge about human colour vision (or about an appropriately chosen part thereof). If so, premise P1 should be accepted as an appropriate description of a legitimate

thought experiment". (Note that Nida-Rümelin's remarks here presuppose not just that it is in-principle possible to have complete physical knowledge about human color vision, but more strongly that Mary could have such knowledge prior to her release from the black-and-white room.) A good case can be made for denying P1, however, as per what I hereby dub the 'Incomplete Physical Knowledge' strategy.

To start, note that nearly all participants in the debate over the import of the knowledge arguments take for granted that physical knowledge does not include knowledge of subjective or qualitative aspects of consciousness. Hence Nagel links physical knowledge to physical theory, and takes the latter to concern only "objective" phenomena, such that it would be "impossible" for physics to accommodate a subjective point of view, and Jackson motivates P1 on grounds that "it is plausible that lectures over black-and-white television might in principle tell Mary everything in the physicalist's story. You do not need color television to learn physics or functionalist psychology" (295). Opponents of the knowledge argument similarly take for granted that complete physical knowledge fails to be knowledge of any subjective or qualitative aspects of reality there might be. This is true on the 'Ability Hypothesis', according to which what Mary gains upon leaving her room is a new ability rather than a new piece of knowledge (see, e.g., Lewis 1990, Nemirow 2006). It is true on what Nida-Rümelin calls the 'Complete Physical Knowledge without Knowledge of all the Physical Facts' strategy, which proceeds by accepting P1 but denying the inference from P1 to C1 (according to which Mary knows all the physical facts about human color vision before her release), on grounds that since physical knowledge is not of subjective or qualitative facts, and since some physical facts are subjective or qualitative, Mary's having complete physical knowledge does not entail her knowing all the physical facts (see, e.g., Harman 1990, Flanagan 1992, and Alter 1998). It is true on the 'Acquaintance Hypothesis', which again grants that Mary has complete physical knowledge, and moreover knows all the physical facts, prior to leaving her room, but maintains that what she gains upon leaving the room is 'knowledge by acquaintance', which, in not being 'informational', doesn't threaten physicalism. And it is true on the popular 'two ways' or what Nida-Rümelin calls the 'New Knowledge; Old Fact' approach, according to which, while Mary does gain new knowledge upon leaving her room, this is simply knowledge about a different, qualitatively informed way of thinking about a fact she already knew, such that the supposition in P1 that prior to leaving her room Mary had complete physical knowledge, sufficient unto knowing all the physical facts, is not undermined.

But need the physicalist agree that physical knowledge is 'objective' in the sense, in particular, of failing to be of any subjective or qualitative aspects of reality there may be? No, for two reasons.

First, such a view is in tension with physicalism. The ontological physicalist of whatever stripe maintains that the lower-level physical goings-on, either individually or in appropriately complex combination, provide a basis for all of natural reality. Given that natural reality includes consciousness and its associated subjective and qualitative aspects, and given that the physicalist denies that 'basic' physical goings-on either have or bestow mentality, the physicalist should correspondingly maintain that some appropriately complex physical goings-on—namely, those that are either identical with (on a reductive physicalist view) or which realize (on a nonreductive physicalist view) these aspects of consciousness—constitutively involve the instantiation of subjective and/or qualitative features which (again, given the supposed truth of physicalism) are themselves physical or physically acceptable features. The questions of whether these features should be seen as had by the highly complex physical micro-configurations that accompany the instantiation of the features, and of whether, if so, the associated micro-configurations should be taken to have a first-person perspective, represent choice points: what answers are given to these questions will depend on further details about the form of physicalism at issue, as well as what account is given of the creatures having such conscious states. However the physicalist answers these questions, they are in any case committed to complex lower-level physical goings-on's involving the instantiation of lower-level physical features which, in being either identical to or such as to realize subjective and qualitative aspects of natural reality, are themselves properly deemed subjective and qualitative.

Now, complete physical knowledge is, as a matter of definition, complete knowledge of any and all actual and/or potential physical goings-on, including knowledge of the potential instantiation of any subjective and qualitative physical features which suffice, either by way of identity (on a reductive view) or realization (on a nonreductive view), for any subjective and qualitative aspects of consciousness there might be.[8] How is one to gain knowledge of subjective and qualitative physical features? The physicalist can and should maintain, plausibly enough and without any clear threat to their position, that such knowledge can only be gained by acquaintance—that is, by being in position to experience these subjective and qualitative features either directly (on a reductive physicalist view) or indirectly (on a nonreductive physicalist view). Consequently, the physicalist can and should deny (in particular) that Mary has complete physical knowledge about human color vision before her release—that is, they can and should deny P1.

[8] Nothing in this claim turns on whether knowledge has to be broadly propositional—i.e., of states of affairs as opposed to properties or other features—since knowledge of (the potential instantiation) of properties can be translated, if so desired, into propositional terms (as, e.g., knowledge that a certain property would be instantiated under such-and-such circumstances).

I will shortly offer a diagnosis of why, in spite of its being in clear tension with the truth of physicalism, P1 has been taken for granted even by opponents of the knowledge argument. First, though, it's worth noting that the Incomplete Physical Knowledge strategy improves on certain alternative strategies of response to the knowledge argument, while arguably preserving the insight common to many of these responses—namely, that the fact that certain kinds of experience are prerequisite to grasping certain features of reality is no barrier to the physical acceptability of these features. Consider, for example, Conee's (1994) case for taking the Acquaintance hypothesis to block the physical unacceptability of phenomenal qualities:

> The physical facts may include every fact about qualia. Still, the physical story does 'leave out the qualia', in the sense that knowledge of the physical facts does not imply knowledge of the qualia. Gaining knowledge of phenomenal qualities, though, is no more than a matter of making their acquaintance by attentive experience. It requires only entering into a new cognitive relation to the qualities, not learning any new information. It gives us no reason to doubt that everything is physical. (148)

As Nida-Rümelin points out, a proponent of the knowledge argument is likely to respond that if the having of an experience with the qualitative physical property (*Q*) in question is required for knowledge of it by acquaintance, such acquaintance is moreover "a necessary condition for being able to know (in the relevant sense) that an experience has *Q*". What has not been previously appreciated is that the physicalist can and arguably should simply grant that acquaintance is a necessary condition for knowing certain physical facts—namely, those providing a subjective or qualitative basis for any subjective or qualitative aspects of consciousness there may be. And similarly for the usual physicalist responses to the Ability hypothesis, the 'two-ways' strategy, and the 'Complete Physical Knowledge without Knowledge of all the Physical Facts' strategy. Rather than stand opposed to the intuitively compelling take on Mary's reaction to seeing a ripe tomato—'so *this* is what it is like to see red!'—according to which she does come to know a new fact after leaving her room, the physicalist can rather simply agree that Mary gains new knowledge, and then 'modus tollens' the anti-physicalist conclusion as rather showing that prior to leaving her room, Mary didn't have complete physical knowledge (of human color vision, in particular), after all.

Why has this strategy of response been previously neglected? I diagnose this, and the associated acceptance of P1 by opponents and proponents of the knowledge argument alike, as reflecting a mistaken characterization of the physical goings-on—one which is (a) overly representational, (b) overly restricted, and (c) problematically qualitatively etiolated.

To start, recall that the motivation for P1 reflects the supposition that physics does not contain reference to subjective or qualitative phenomena. Hence it was that Nagel claimed that physical 'theory' concerns only 'objective' goings-on, and that Jackson claimed that qualitative experience is not needed to learn physics.

The first problem with this motivation for P1 is that it characterizes the physical basis in overly representational terms. Yes, physicalists look to physics as a guide to the compositionally basic goings-on, but representation (not to mention theory) is one thing, reality another. Recall the operative conception of the physical goings-on, according to which these are the goings-on that are treated, approximately accurately, by present or (in the limit of inquiry, ideal) physics (see Ch. 1, §1.4.1; Wilson 2006). On this and other physics-based conceptions of the physical, physics gives us a working handle on the compositionally basic goings-on, but even in the limit of inquiry, there is no supposition that physical theory will be up to the task of offering a complete description of these goings-on.

The second problem with this motivation is that (and even bracketing the previous point) it presupposes an overly restricted characterization of the physical goings-on. As discussed in Ch. 1 (§1.4.2) and elsewhere, the goings-on explicitly referenced in physics provide the starting point, not the end point, of the lower-level physical base. To be sure, the immediate targets of physical theory—e.g., subatomic particles or their field-theoretic or wave-functional correlates—do not individually instantiate or otherwise encode subjective or qualitative aspects of consciousness of the sort we enjoy. But that doesn't show that sufficiently complex physical goings-on do not do so, any more than the fact that the goings-on explicitly referred to in physical theory do not individually instantiate or otherwise encode thermodynamic processes and properties shows that sufficiently complex physical goings-on do not do so.

A third problem might be thought to reflect a somewhat more substantive motivation for P1—namely, that (so the story usually goes) insofar as physics is concerned only with the trajectories and other behaviors of the goings-on that are the immediate target of physical theory, and insofar as such behaviors and associated laws are completely described in structural, non-qualitative terms, no amount of complex or even causal combination could ever eventuate in the instantiation of physical properties up to the task, so to speak, of either being or serving as a physical basis for such qualitative aspects of consciousness. Structure, one might say, can only beget more structure.

Now, this is a more substantive motivation for P1, but drawing upon the previous responses, the physicalist can reasonably maintain that it mischaracterizes the physical goings-on, even at low orders of complexity, as completely qualitatively etiolated.

Such an abstract, completely structuralist conception of the compositionally basic entities is not forced by physical theory. On the contrary: it is common for physicists to speak of particles or other lower-level physical phenomena as 'feeling' forces, where the reference here might be thought to be both qualitative and sensitive to perspective—the force coming from this direction rather than that, and applied to the entity in its location in ways that then enter into its moving in such-and-such a way. This degree of qualitativity, in which a noncomplex physical entity is capable of registering a local perspective and associated sensitivity to the environment, in a way that is not entirely qualitatively etiolated, clearly need not be seen as involving any measure of consciousness or mentality, even of a 'protopsychic' variety. On the contrary, it reflects a naturalistic point of view according to which physical goings-on are real and substantial, and when they move, they do so not because, e.g., they must play a predetermined part in the harmony of the spheres, but rather for some salient reason or reasons—for instance, the influence of one or more felt forces or interactions. This understanding of noncomplex physical goings-on suggests that we should distinguish qualitativity and consciousness: sometimes these go together, as in creatures like us; but qualitativity is a weaker notion, that may be present, e.g., in the response of a particle to the forces acting on it, or more generally, in a particle's engaging in interactions in ways reflecting sensitivity to the relevant environmental circumstances.

To be sure, from recognition of this amount of qualitativity in the physical goings-on it's still a considerable leap to get to the instantiation of subjective or qualitative aspects of consciousness. But given that some explanatory gaps (e.g., in the case of complex nonlinear systems) do not have anti-physicalist import, the burden is now on the anti-physicalist to establish that the seeming explanatory gap between physical goings-on and subjective and qualitative aspects of conscious experience must be taken to have such import. The strategy of the knowledge arguments crucially proceeds by way of P1 and the associated characterization of complete physical knowledge, as failing to be knowledge of any subjective or qualitative aspects of consciousness there may be. But the physicalist can and should reject this characterization, both as effectively begging the question against their view (since for the physicalist, complete physical knowledge must include knowledge of those physical features that are identical with or serve as a subjective or qualitative basis for such aspects of consciousness), but also on grounds that there's no independent reason to accept such a characterization.

I conclude that the knowledge arguments do not provide compelling reason to think that consciousness and its associated subjective and qualitative aspects are actually physically unacceptable, much less actually Strongly emergent.

7.1.2 The conceivability argument

I turn now to the conceivability argument advanced and developed by Chalmers (1996, 1999, 2009, and elsewhere), according to which the conceivability of zombies—creatures which are functional and physical duplicates of creatures like us, but which are lacking in any conscious mentality—is taken, in combination with certain other commitments, to establish the Strong emergence of consciousness.

Like the knowledge arguments, the conceivability argument relies on the presence of an explanatory gap, as needed to make room for the conceiving in question. As Levine (1983) describes the connection:

> If there is nothing we can determine about C-fiber firing that explains why having one's C-fibers fire has the qualitative character that it does—or, to put it another way, if what it's particularly like to have one's C-fibers fire is not explained, or made intelligible, by understanding the physical or functional properties of C-fiber firings—it immediately becomes imaginable that there be C-fiber firings without the feeling of pain, and *vice versa*. We don't have the corresponding intuition in the case of heat and the motion of molecules—once we get clear about the right way to characterize what we imagine—because whatever there is to explain about heat is explained by its being the motion of molecules. (359)

Chalmers's argument goes beyond the knowledge arguments, however. To start, his argument relies on the conceivability of zombies, as opposed to the mere absence of explanation or associated failures of knowledge. That said, an appeal to conceivability alone isn't much of an advance on the previous arguments, since just as physicalists need not deny that there are explanatory gaps between consciousness and lower-level physical goings-on, neither need physicalists deny that zombies are conceivable, as long as they are not forced to take such a broadly epistemic fact to have metaphysical import (in particular, of the anti-physicalist variety).

The primary advance of Chalmers's argument rather lies in his situating the conceivability of zombies in an independently motivated framework—'epistemic two-dimensionalism', or E2D, to be explicated in more detail shortly—according to which certain facts about meaning, which are taken to be a priori accessible, can be used to identify or establish certain facts about modality, expressing or encoding what is genuinely metaphysically possible (necessary, contingent, impossible). It is commonly assumed that the mode of a priori access to meanings that enters into the E2D strategy proceeds by way of conceiving. Consequently, commitment to the E2D strategy for gaining access to modal truth, and to this strategy's

being implemented by means of a conceiving-based epistemology of meanings, provides an independent basis for taking the conceivability of zombies to have anti-physicalist metaphysical import, as a case in point of a systematic connection between conceivability and metaphysical (as opposed to merely epistemic) possibility.

Reflecting the crucial role that E2D plays in his argument, Chalmers has come to call his argument, more accurately (and again, for reasons that will be made clear shortly), 'the two-dimensionalist argument against materialism'. The argument proceeds essentially as follows:

1. It is conceivable that there is a world which is physically exactly like our world, but in which there is no consciousness.
2. If the world described in (1) is conceivable, then it is metaphysically possible. (E2D)
3. If the world described in (1) is metaphysically possible, then physicalism is false.
4. Therefore, physicalism is false.
5. In particular, consciousness is physically unacceptable (and moreover might be Strongly emergent).

In aiming to provide independent motivation for taking explanatory gaps to have anti-physicalist metaphysical import, Chalmers's argument is different from, and to my mind improves on, the knowledge arguments. Nonetheless, as I'll presently argue, Chalmers's line of thought is also ultimately unsuccessful, and hence fails to motivate taking consciousness to be physically unacceptable, much less Strongly emergent.

The focus of my critical attention is on the second premise.[9] I'll first present Chalmers's two-dimensionalist argument in more detail, highlighting the attractiveness of the E2D strategy that lies at the argument's core. Next, drawing on Biggs

[9] Rejecting the first premise seems likely to lead directly to stalemate (a problem discussed in Biggs and Wilson 2019 and 2020), not just because what counts as 'conceiving' is unclear, but because (even holding fixed a given understanding of the conceivability at issue) people disagree about what they conceive. (Similar remarks apply to variations on the theme of conceiving as involving, e.g., rational intuition as per Bealer 2002.) Rejecting the third premise is ultimately indecisive. One might deny that the metaphysical possibility of a zombie world suffices to establish the falsity of physicalism, on grounds that the premise presupposes that any relation between conscious states and lower-level physical states compatible with physicalism must be one according to which the latter metaphysically necessitate the former, but (the objector maintains) this presupposition is incorrect. Recall, e.g., that it would suffice for the physical acceptability of conscious states that the latter be functionally realized by physical states, in such a way that the powers of the former are, on any given occasion, a proper subset of those of the latter; but if powers are only contingently associated with features, then the physical states that realize conscious states in worlds with our laws of nature might not do so in worlds with different laws of nature—compatible with, in particular, the existence of zombie worlds. This response to Chalmers's two-dimensional argument is ultimately indecisive, however, since he might maintain that the argument applies even when the holding of the actual physical laws is built into the specification of worlds in which the lower-level physical base goings-on are present.

and Wilson (2019 and 2020), I'll suggest that there is an alternative, and superior, way in which the E2D strategy might be implemented—namely, by appeal to an abduction-based rather a conceiving-based epistemology of the meanings entering into this strategy. I'll then argue that it is far from clear that the genuine possibility of zombies, or the associated Strong emergence of consciousness, is output from E2D, when this framework is implemented using abduction rather than conceiving. One might wonder, as against this line of thought, whether abduction would be suited for purposes of implementing E2D, given that (as above) the access to the meanings which are in turn supposed to provide a basis for access to modal truths is supposed to proceed in a priori fashion. Here again, I draw on joint work with Biggs (Biggs and Wilson 2017), where we argue that, contra common assumption, abduction is an a priori mode of inference—as a priori as conceiving, in particular. The upshot at the end of the day will be that, like the knowledge arguments, Chalmers's two-dimensional argument fails to establish that consciousness is actually physically unacceptable, much less Strongly emergent. That's the overview; now for the details.

Epistemic two-dimensionalism

What is missing from the knowledge arguments is independent good reason to take the explanatory gaps that they highlight to have anti-physicalist metaphysical import. Chalmers's intended route to such good reason takes as its starting point Kripke's (1972/80) undermining of the traditional supposition that modal truths are (always) a priori accessible. Traditionally, claims about what is possible or necessary were taken to be knowable a priori, reflecting that such modal claims were presumed not to hinge on actual facts. Kripke argues, however, that many necessary truths about individuals and natural kinds can be known only a posteriori. For example, one could not have determined a priori that water is identical to H_2O; that water is H_2O is an empirical matter. Nonetheless, Kripke plausibly asserts, given that water is H_2O, then water is necessarily H_2O; so, that water is necessarily H_2O is knowable only a posteriori. For another example, one could not have determined a priori that this lectern is made of wood; that a certain lectern is made of wood is an empirical matter. Nonetheless, Kripke plausibly maintains, given that a certain lectern is made of wood, it follows that it is not possible for that very lectern to be made of some completely different kind of material—again, a modal claim knowable only a posteriori. More generally, Kripke's results might be seen as undercutting the prospects of our having much, if any, a priori modal knowledge about broadly scientific goings-on.

These consequences would be undesirable, not just for purposes of assessing the truth of physicalism, but more generally because much theory and practice (philosophical, legal, semantic, scientific, etc.) presupposes that we can know modal truths independently of, or at least prior to the end of, empirical inquiry. We might thus hope that, notwithstanding Kripke's results, the traditional link

between necessity and a priority could be restored, to at least some considerable extent.

This is the promise of epistemic two-dimensional semantics (E2D), advocated by Chalmers (1996, 2006*b*, 2009), Chalmers and Jackson (2001), and others.[10] E2D can be heuristically seen as refining and generalizing Frege's (1892/2010) suggestion that there are two kinds of meaning: sense and reference. On the Fregean picture, sense represents an aspect of meaning that is a priori accessible to a competent speaker, whereas reference represents an aspect of meaning that may fail to be so accessible. For example, Frege maintained that a competent speaker could know a priori the senses of 'Hesperus' and 'Phosphorus' as 'the first star seen in the evening' and 'the first star seen in the morning', respectively, while not knowing the empirical fact that these expressions have the same referent (namely, the planet Venus). Moreover, on a standard reading of Frege's view, our a priori access to sense provides a basis for a priori knowledge more generally, in that whether a claim is knowable a priori depends entirely on the nature of the relations among the senses of its constituent expressions and the nature of our access to them. For example, on Frege's view 'bachelors are male' is knowable a priori because the senses of 'bachelor' and 'male' appropriately overlap, and one can grasp the senses of 'bachelor' and 'male' in a way revealing this overlap.

E2D similarly aims to characterize two distinct aspects of meaning, each represented as 'intensions': functions from possible worlds to extensions. More carefully, and following Chalmers, intensions are functions from either scenarios (centered worlds, incorporating a subject's perspective) or scenario-world pairs to extensions. The 'primary' intension of an expression E takes each scenario s to the extension of E in s on the supposition that s is actual. The 'secondary' intension takes each scenario-world pair $< s, w >$ to the extension of E in w given that s is actual. The E2D strategy supposes that many natural kind expressions of the sort entering into a posteriori necessities are associated with primary and secondary intensions, so understood. (Note that the exposition to follow is abbreviated, owing to considerations of space. The interested reader is directed to Chalmers 2006*b* and 2009 and Biggs and Wilson 2019 and 2020 for more detailed expositions.)

There are several useful ways of characterizing these intensions. On one, primary intensions track extensions in contexts of utterance, and secondary intensions track extensions in contexts of evaluation given a context of utterance. On another, primary intensions track extensions in worlds 'considered as actual', and secondary intensions track extensions in worlds 'considered as counterfactual', holding fixed which world is actual. In any case, the promise of E2D ultimately

[10] See, e.g., Peacocke 1993, Chalmers 1996, Boghossian 1996, and Gertler 2002.

lies in the suggestion that primary and secondary intensions are connected in such a way as to provide a basis for a priori knowledge of a wide range of modal truths. To understand the suggested connection, it is useful to think in terms of a two-dimensional matrix, with worlds 'considered as actual' along the vertical, and worlds 'considered as counterfactual' along the horizontal (cf. Chalmers 2006*b*, following Stalnaker 1999). For example, the two-dimensional matrix for the intensions associated with the term 'water' would look something like Figure 7.1:

'water'	H_2O-world	XYZ-world	…
H_2O-scenario	H_2O	H_2O	…
XYZ-scenario	XYZ	XYZ	…
…	…	…	…

Figure 7.1 Two-dimensional matrix of intensions for 'water'

The entries in the matrix encode that our access to appropriate intensions, on the E2D strategy, provides a basis for our knowing a priori that *if* water (or the watery stuff) is actually H_2O, *then* water is necessarily H_2O; and *if* water (or the watery stuff) is actually XYZ, *then* water is necessarily XYZ; and so on. Correspondingly, while it is (as per Kripke's results) a posteriori that water is necessarily H_2O, the a posteriori contribution to this and related necessity claims concerning natural kind terms is limited to discharging the antecedent of a conditional known a priori—compatible with our having a priori access to a great deal of modal knowledge.

Granting that we have independent reason to accept the E2D framework and associated strategy for reforging the link between a priority and modality, the question remains which epistemology of intensions (meanings) should be taken to be operative in implementing this strategy. As prefigured at the start of this section, it has commonly been taken for granted that the epistemology of modality must rely on conceiving. Hence Gertler (2006) says:

> Conceivability is the only guide to necessity; our concepts, and the intuitions about possibility that derive from them, provide our only grip on modal claims. […] Since a claim of impossibility cannot be established by considering the actual world alone (though of course it can be refuted in this way), [one] must consider whether certain non-actual scenarios are possible. And the only way to determine this is to use the method of conceivability. (205)

Given a conceiving-based epistemology of intensions (CEI), we are close to the anti-physicalist conclusion of Chalmers's conceivability argument. CEI implies that 'zombie' has a positive extension at some world if one can conceive of a zombie

world; E2D coupled with CEI implies that, if one can conceive of a zombie world, then zombies are metaphysically possible. But as per the first premise of Chalmers's argument (which I am granting), one can conceive of a zombie world. So zombies are metaphysically possible, which given the third premise of Chalmers's argument (which premise I am also granting) implies that consciousness is not physical, and moreover might be Strongly emergent.

This is an interesting line of thought, but as I'll now argue, we have both negative and positive reason to reject it.

Conceiving-based vs. abduction-based epistemologies of intensions

Let's start by asking: why think that the only epistemic access to the intensions of our concepts or terms proceeds via conceiving, as per CEI? Why not, in particular, allow that such access might proceed via inference to the best explanation, or abduction, as per an abduction-based epistemology of intensions (AEI)? Why not allow that various theoretical desiderata or abductive principles—plausibility, compatibility with other beliefs, unifying power, ability to resolve certain problems, explanatory fruitfulness, ontological and ideological parsimony, and the like—can come into play in identifying the extensions of certain concepts or terms in non-actual scenarios (which extensions enter into constituting the associated intensions)? After all, on the face of it such considerations might well be relevant.

Plausibly, AEI has not been taken seriously as a basis for implementing the E2D strategy on grounds that, if the strategy is to do the job of reforging an a priori accessible link between meaning and modality, the operative epistemology of intensions must be a priori (involve an a priori mode of inference); and while conceiving is a priori, abduction is not. As Biggs and I argue in our (2017 and 2019), however, abduction *is* a priori. In our papers, we offer many positive and defensive motivations in support of taking abduction to be a priori. Here I offer sketches of two argumentative lines of thought for this position, and call attention to some historical and contemporary precursors of our view.

First, those taking abduction to be a posteriori typically do so on grounds that whether abduction has epistemic value, sufficient unto conferring justification on conclusions output from that mode of inference, depends on contingent facts. As Beebe (2009) notes:

> [P]ractically everyone who works on abductive inference believes that such inferences are justified empirically and that the theoretical virtues are broadly empirical and contingent marks of truth. (625)

What motivates this understanding of the epistemic value of abduction? To fix ideas, consider a specific abductive principle—say, *Parsimony*, a version of ontological parsimony according to which, *ceteris paribus*, one should choose

the theory involving the fewest fundamental type-level ontological commitments. Why think that the epistemic value of *Parsimony* (hence abduction, as partly constituted by *Parsimony*), depends on experience? Something like the following sort of case seems to be at issue:

> Sam gains a bunch of evidence about the world and identifies two theories, T and T^*, that would appropriately accommodate the evidence. At t_1, she picks T as the most parsimonious theory, but later, at t_2, gains new evidence supporting T^*. So choosing T on the basis of *Parsimony* led her astray.

As Biggs and I argue, however, such a case (or series of cases, or proportion of such cases) at a world does not show that abduction lacks epistemic value in that world, but rather that the supposition that other things were equal was, as it turned out, mistaken.[11] As we put it in our (2017):

> Consider how the objector would describe cases in which, they claim, *Parsimony* is taken to initially support a theory that further evidence later disconfirms. According to the objector, at t_1 *Parsimony* supports theory T over theory T^* given explananda E, and at t_2 we discover additional explananda E', which T^* explains better than T does. [...] A more accurate description of such cases is as follows: at t_1 we mistakenly believe that T and T^* are otherwise equal (because we are unaware of E'), but at t_2 we discover our mistake. Accordingly, rather than the discovery of E' disconfirming a theory initially supported by *Parsimony*, that discovery reveals that we initially applied *Parsimony* too hastily. And similarly, mutatis mutandis, for any cases that might be thought to figure in empirically [disconfirming] *Parsimony*. (745)

Generalizing, we suggest that the usual supposition that abduction is a posteriori reflects a failure to appreciate that the *ceteris paribus* clauses associated with abductive principles operate to shield them from empirical disconfirmation.

By way of support for this line of thought, note that it is commonly recognized that failure to possess all the relevant data can lead to false belief via modes of inference whose epistemic value is taken to be a priori. As B. Russell (2017) recently observed:

> You might initially be a priori justified in believing that no matter how happiness has been produced it is intrinsically good [...], or that it is always wrong to punish an innocent person [...], and later think of counterexamples to such claims (e.g., happiness had through the suffering of others or punishing an innocent person to prevent some evil men from punishing him and many other innocent people).

[11] As I'll discuss shortly, this is not to say that the initial application of *Parsimony* was inapt.

For another example: were one to initially apply modus ponens to some premises later revealed to be false, that would show not that modus ponens is a posteriori, but rather that one was working with faulty (incomplete or inadequate) data. Similarly, we maintain, for the case of Sam's abductive inference: that new evidence might undermine the conclusion of a given abductive inference goes nowhere towards establishing that abduction is a posteriori.

To be sure, there are further choice points associated with cases in which one applies a given mode of inference to faulty (incomplete or inaccurate) data. As above, B. Russell (2017) supposes that in such cases the conclusions, even if false, may be a priori justified (such that, notwithstanding that Sam was in fact mistaken in thinking that all else was equal between T and T^*, she was nonetheless justified in taking T to be the best explanation at t_1). By way of contrast, Chalmers builds into his (2002) account of conceivability as an a priori guide to possibility that the conceivers are suitably ideal and in possession of all relevant information, and that the products of such conceivings are indefeasible, which suggests that the conclusions of non-ideal, not-fully-informed conceivings might not be a priori justified by Chalmers's lights (such that Sam was not a priori justified in taking T to be the best explanation at t_1). Another choice point might reflect the distinction between internalist and externalist understandings of the justification in these cases. For present purposes, what is important is that cases such as that involving Sam do not in themselves provide a basis for taking abduction to be a posteriori. On the contrary, such cases are reasonably interpreted as indicating, as above, that insofar as the *ceteris paribus* clauses associated with abductive principles operate to shield them from empirical disconfirmation, the epistemic value of these principles (and of abduction more generally) is a priori, with failures of abductive inference to lead to true conclusions being due not to any purported absence of epistemic value of abductive principles or abduction, but rather to failure of the abductor to be in possession of all the relevant data.[12]

A second line of thought in favor of taking abduction to be a priori proceeds by attention to the roles experience may play in inferential reasoning. Consider a claim p. There are four ways in which experience might play a role in the course of a particular belief in p coming to be justified. Experience might play a role in

1. acquiring the concepts required to entertain p,
2. acquiring the evidence required to justify belief in p,
3. justifying belief in the epistemic value of the mode of inference used to justify belief in p, or
4. acquiring or learning to deploy the mode of inference used to justify belief in p.

[12] There is more to say by way of filling in and defending this line of thought; I direct the interested reader to Biggs and Wilson 2017 and 2019.

Can a belief for which experience plays an ineliminable role along one or more of (1)–(4) be justified ('entirely') a priori? As Biggs and I discuss in our (2017), it depends on which role is at issue. Re (1): it is commonly maintained that belief in *p* can be justified a priori even if experience is needed to acquire the concepts required to entertain *p*. For example, belief in 'sisters are siblings' can be justified a priori even if we need experience to acquire the concepts or meanings of 'sister' and 'sibling'. Re (2): it is commonly maintained that belief in *p* cannot be justified a priori if experience is (at all) needed to acquire the evidence supporting *p*. This is the sense in which any reliance on experience or empirical facts suffices to render the associated justification a posteriori. For example, belief in 'water is H_2O' is commonly taken to be justified a posteriori, on grounds that justifying this belief requires, among other things, acquiring empirical evidence to the effect that water and H_2O are spatiotemporally coextensive. Re (3): it is commonly maintained that belief in *p* cannot be justified a priori if experience is needed to justify belief in the epistemic value of the mode of inference required to justify belief in *p*. Re (4): although the role of experience in an agent's acquiring or learning how to deploy a given mode of inference is not much discussed, it seems reasonable to maintain that this role is relevantly similar to that of (1): in each case, experience contributes to belief (or supposition) formation, not to justification per se. For example, students often need encouragement to think in an appropriately imaginative way about what is possible, but that such imaginative 'training up' is required in order to engage in suitably competent conceiving is not taken to undermine the status of conceiving as an a priori mode of inference. Hence it seems reasonable to maintain that belief in *p* can be justified a priori even if experience is required in order to acquire or learn how to deploy the inferential mode at issue. Correspondingly, among the roles that experience may play in inferential reasoning, only (2) and (3) are such that the playing of this role in the justification of a given belief would render that justification a posteriori.

Now, suppose that belief in a given claim *p* is justified via abduction. Following the considerations just canvassed, whether that belief is justified a priori turns on whether experience enters into its justification via either role (2) or role (3)—that is, on whether experience provides evidence for *p* without which that belief in *p* would not be justified, or whether experience plays (played) a role in justifying belief in the claim that abduction has epistemic value.

Following the discussion above, experience does not enter into abductively justifying the belief in *p* via role (3): the epistemic value of abduction does not rely on experience. Does abductive justification require that experience enter via role (2)? And in cases where experience does enter via role (2) in the abductive justification of a given belief, should the contribution of abduction to this justification be considered a posteriori? No, and no. To start, notwithstanding that abduction can and often does operate on empirical evidence (e.g., 'water and H_2O

are spatiotemporally coincident') to produce a claim (e.g., 'water is H_2O') that is a posteriori justified, the underlying inferential transition can be encoded in conditionals—e.g., 'if water and H_2O are spatiotemporally coincident, then water is H_2O'—which are justified by means of a hypothetical form of abduction. Such a hypothetical abductive inference would be akin to suppositional reasoning in conditional proof, allowing one to identify what would be the best explanation of the state of affairs expressed in the antecedent, were this state of affairs to obtain (were the antecedent to be true). And since such conditional beliefs may be justified without the antecedent claim's being either believed or true, abduction in such cases operates independently of any claim that is justified through experience. Accordingly (and given that the epistemic value of abduction does not rely on experience, as we have argued), such abductively justified conditional beliefs are 'entirely' a priori. Hawthorne (2002, 252) makes a similar point, suggesting that abduction can deliver a priori justification for belief in a conditional whose antecedent describes an 'experiential life history' and whose consequent is whichever theory best explains some aspect of that life history; see also Cohen 2010, 152–3 and Wedgwood 2013.

It follows that abduction is an a priori mode of inference, even when it operates on (non-hypothetical) empirical evidence (registering, e.g., the actual spatiotemporal coincidence of water and H_2O). To be sure, when abduction operates on empirical evidence, it does not produce beliefs with (completely) a priori justification, any more than (e.g.) modus ponens does. But the contribution of empirical evidence in such cases of abductive inference is simply to discharge the antecedent of a conditional whose justification is entirely a priori. The same is true of conceiving; and indeed, it is precisely the appreciation that we can have a priori access to relevant conditionals that serves as the basis of the E2D strategy for regaining, post-Kripke, significant a priori access to modal claims.

Finally, it is worth noting that the view that abduction is a priori has clear historical precursors in the work of Kant and Carnap. In brief (see Biggs and Wilson 2019 for further discussion) Kant took synthetic a priori truths (including most philosophical claims) to be justified via an ampliative mode of inference—plausibly, abduction; Carnap took knowledge of conceptual content to proceed via explicitly abductive 'explication'. More recently, and in addition to Hawthorne (2002), Cohen (2010), and Wedgwood (2013), several philosophers, including Bonjour (1998), Swinburne (2001), and Peacocke (2003), have offered reasons (different from ours; see the comparative discussion in Biggs and Wilson 2017) aiming to establish the a priority of certain ampliative modes of inference or abductive principles.

Summing up: although it has commonly been taken for granted that abduction is a posteriori, there are cases to be made that abduction is a priori; and this view is not as unusual as it might first appear.

I next want to highlight the primary independent reason for thinking that, given that abduction is a priori, we should look to abduction rather than conceiving as a guide to the intensions at issue in the E2D strategy.[13]

This reason reflects what might be called 'access-based' objections to E2D, according to which the common presence of indeterminacy or inconsistency in our concepts indicates that we commonly fail to have a priori access to the intensions of natural kind expressions (among others), to such an extent that the prospects of an E2D-led reforging of the link between meaning and modality are fatally undermined. As Biggs and I argue in our (2020), these access-based objections are not to E2D per se, but rather to E2D when implemented using CEI; for while conceiving alone does not have the resources to resolve indeterminacy or inconsistency in our concepts, abduction *does* have such resources. More generally, Biggs and I argue that when E2D is combined with AEI rather than CEI, the full range of access-based objections can be satisfactorily answered. In what follows, I discuss as an illustrative case in point one salient access-based objection. Afterwards I return to Chalmers's conceivability argument, and the question of whether an implementation of E2D provides reason to think that zombies are genuinely metaphysically possible.

According to the objection at issue, widespread conceptual indeterminacy—better: underdetermination[14]—undermines the prospects for E2D. The line of thought here is that implementing the E2D strategy requires that extensions of expressions (e.g., natural kind terms) have a priori-accessible extensions in every possible scenario; but in cases of conceptual underdetermination the associated expressions lack fixed antecedent extensions in some scenarios. Moreover, the concern runs, such underdetermination is widespread. That there is widespread conceptual underdetermination is suggested by M. Wilson's (1982 and 2006*b*) arguments to the effect that applications of natural kind predicates can and frequently do depend on arbitrary factors. For example, in one of Wilson's toy cases, whether members of an isolated tribe take airplanes to be in the extension of their predicate 'bird' (or linguistic correlate) might depend on whether the first airplane they encounter is overhead (in which case, airplanes are judged to be birds) or is rather on the ground (in which case, airplanes are judged not to be birds). If such historical accident might partly determine whether 'bird' applies to airplanes, then, Wilson reasonably assumes, the full range of the predicate's intension was not antecedently determined. And, Wilson suggests, such underdetermination is

[13] For a detailed and more general comparison of abductive vs. conceiving-based epistemologies of modality, see Biggs and Wilson 2019.

[14] Though the objection and responses have been pitched as involving conceptual 'indeterminacy', 'underdetermination' is a better term for the phenomenon at issue, since 'indeterminacy' has connotations associated more specifically with vagueness or metaphysical indeterminacy, of the sort discussed in the previous chapter.

characteristic of natural kind predicates more generally, undermining hopes that E2D can regain much modal knowledge about such kinds.

Chalmers grants that there may be underdetermination of the sort at issue in Wilson-style cases. In his (1996), he says "There may of course be borderline cases in which it's indeterminate whether a concept would refer to a certain object if a given world turned out to be actual" (364), and more recently, Chalmers grants that it is plausible that "later extensions [of the kind of expressions at issue in Wilson's cases] depend on idiosyncratic developments, and verdicts about such cases are not determinately prefigured in a user's original use of an expression" (Chalmers 2012, 231). Chalmers maintains, however, that

> [B]orderline cases [are] no problem: we can allow indeterminacies in a primary intension, as we sometimes allow indeterminacies in reference in our own world. (1996, 364)

But as Biggs and I note, while Chalmers is right that E2D can tolerate some conceptual underdetermination, it cannot tolerate widespread underdetermination, on pain of undermining E2D's *raison d'être* of providing an appropriately expansive basis for our a priori knowledge of a wide range of modal truths, including conditional such knowledge of necessary a posteriori truths (such that, e.g., we can know a priori that *if* the actual world is one where water is H_2O, then it is necessary that water is H_2O). As such, widespread conceptual underdetermination would undermine E2D—at least if such underdetermination is insuperable.

Might Chalmers maintain either that conceptual underdetermination is not really widespread, or that in any case such underdetermination is superable using the resources of conceiving? Neither strategy is promising.

First, conceptual underdetermination is widespread, for both natural kind and other predicates (more generally, expressions). Besides the Wilson-style cases, Biggs and I note two further sources of conceptual underdetermination. The first concerns cases of vague predicates of the sort associated with Sorites sequences (e.g., 'red', 'bald', 'rich', 'cell', 'part', etc.). As Raffman (1994) compellingly argues, the application of such predicates can be and often is typically determined in a given case by arbitrary contextual and/or psychological factors. For example, the breaking point at which one stops (or starts) applying the predicate 'red' in a given color-spectrum sequence may non-systematically depend on where along the spectrum one starts (as per the phenomenon of hysteresis) as well as on arbitrary psychological factors (fatigue, boredom, etc.). For Wilson-style reasons, the dependence of extensions on arbitrary historical and other factors supports thinking that the associated expressions do not come with antecedently fixed extensions—i.e., these extensions are to some extent conceptually underdetermined. But those discussing the phenomenon of vagueness often register

its being likely that *most* expressions in ordinary language (including scientific language) are vague—in which case Raffman's results show that most expressions are conceptually underdetermined.

A second source of widespread conceptual underdetermination reflects the historical scientific record, which supplies many cases of underdetermination involving natural kind terms. For example, 'acid' initially was taken to refer to only oxygenated substances, but was later applied to HCL, for theoretical reasons now largely discarded; dispute persists over whether Newtonian uses of 'mass' apply in relativistic contexts; the decision to classify whales as mammals was a controversial affair; the classification of Pluto as a planet, or not, remains a topic of dispute.

It seems clear, then, that conceptual underdetermination is widespread. Hence such underdetermination will be 'no problem' for E2D only if the operative epistemology of intensions (meanings) is capable of resolving the underdetermination. Now, Chalmers (1996) does consider whether conceivers might be able to eliminate underdetermination by foreseeing relevant accidents. In Wilson's toy case, for example, Chalmers says that one "might try to classify these two different scenarios [airplane first seen in the sky or on the ground, respectively] as different ways for the actual world to turn out, and therefore retain a fixed, detailed primary intension" (364). On this strategy, the fully determinate primary intension of 'bird' includes planes in its extension if the tribe members first see a plane overhead, but not if they first see it on the ground. Either way, according to the suggestion, the underdetermination in the tribe's expression 'bird' is resolved. And Chalmers might attempt to implement a similar strategy as a general means of overcoming conceptual underdetermination involving vague or scientific predicates or other expressions.

Such a strategy has traction against the underdetermination objection only if conceivers can foresee how intensions are sensitive to accidental or arbitrary factors. But as Biggs and I see it, a deeper lesson of Wilson's case, as well as of Raffman's discussion, is that the influence of such factors cannot be foreseen, at least not by conceiving alone. Determinism and such aside, those using conceiving alone might apply 'bird' differently even relative to the same 'accident'. After all, there are any number of respects of dissimilarity between airplanes and birds, even when the former are in flight, and a minor difference in attention to these features might result in a different decision about whether 'bird' applies to a flying airplane. And as Raffman points out, the factors entering into the application of a vague predicate are not just arbitrary but are unsystematically so, as is reflected in different applications of vague expressions even against relevantly the same background conditions. We can register, post hoc, extensions resulting from whatever decision was in fact made, but in cases of arbitrary determination there is no way to antecedently identify these extensions and corresponding intensions

through conceiving alone—there is simply no fixed extension to 'conceive', even taking relevant circumstances into account.

Similarly, there is no case to be made that even an idealized conceiver could determine the extension, in every scenario, of natural kind expressions pertaining to 'mass', 'planet', and so on. As a matter of historical fact, the considered extensions of these terms have been heavily informed by abductive considerations, taking into account plausibility, compatibility with other beliefs, unifying power, ability to resolve certain problems, explanatory fruitfulness, ontological and ideological parsimony, and so on. But as Chalmers and others implementing E2D using CEI emphasize, conceiving does not involve these sorts of abductive or ampliative resources.[15] Correspondingly, conceiving alone is simply not up to the task of overcoming conceptual underdetermination in these cases.

By way of contrast, AEI is up to the task of overcoming conceptual underdetermination, in the cases of natural kind expressions and more generally. For when considering whether to apply an expression in a given scenario, abductors can consider not only historical accident and psychological variability, but also any non-demonstrative rational grounds that might push towards one extension rather than another. Hence abduction, unlike conceiving, can be productive (it is ampliative, after all), allowing those who are identifying intensions to consider how the concept *should* be applied, given the usual abductive principles and associated theoretical desiderata, even when the application conditions are antecedently to some extent underdetermined. Since abduction can, in a rational way, go beyond what expressions antecedently encode, AEI has the potential to overcome conceptual underdetermination, extending applications of natural kind expressions to new scenarios or situations, on ultimately rational grounds. For example, an idealized version of a competent user of 'bird' who has never before seen an airplane would plausibly be in position to consider and compare theories of the intensions of 'bird' (as ever including or as always excluding airplanes, however first encountered), by attention to which theory would be, among other desiderata, most plausible, consonant with our other beliefs, unifying (e.g., of experience of flying entities), fruitful, ontologically and ideologically parsimonious, and so on. Similarly for the cases of 'mass', 'Pluto', 'water', and so on. Indeed, in these and many related cases, the antecedent underdetermination has been resolved (when it has been resolved) not through accident, but through abductive reasoning. Here it is worth noting, by the way, that there is no barrier to an idealized abductor engaging in abductive

[15] For example, Chalmers and Jackson (2001) are explicit (when arguing, in particular, against Block and Stalnaker's 1999 claim that the justification for the conditional claims output by E2D might rely on broadly abductive considerations) that in their view abductive considerations do not play any 'essential justificatory' role in the cases at hand; see especially 342–50.

deliberations about how best to extend a given expression or concept, for this is more or less what language-users do as a matter of course.

Similar results attach to other access-based objections (see Biggs and Wilson 2020 for discussion). Even focusing just on underdetermination, however, the upshot is clear: E2D when implemented using a conceiving-based epistemology of intensions faces difficulties to which an abduction-based epistemology of intensions, in having ampliative resources of the sort that conceiving fails to have, is in position to respond.

Revisiting the possibility of zombies

This result in hand, I now return to the status of Chalmers's two-dimensionalist argument against materialism, and in particular, to the question of whether the independently desirable E2D framework supports taking zombies to be metaphysically possible. Effectively, the question is: are there worlds that are physical and functional duplicates of our own (including, let us grant, worlds where the physical laws are the same as our own) in which the intension of 'zombie' has a non-null extension—that is, where there are zombies?

We can start by observing that on AEI, an attempt to answer this question will proceed by attention to any and all relevant considerations, which may include but are not restricted to any modal intuitions we might have. The usual theoretical desiderata will be relevant, as will be considerations such as the desirability of properly accommodating the causal efficacy of consciousness, and (perhaps relatedly) the desirability of conforming, for purposes of systematicity and plausibility, to what our best sciences give us reason to believe. Accordingly, different 'theories' of the intensions at issue will rate better or worse, depending on how well they do at accommodating or satisfying these diverse considerations.

A theory on which the extension of 'zombie' has a non-null extension at some world scores positively in accommodating the seeming conceivability of zombies. On the other hand, insofar as such a non-null extension implies that consciousness is Strongly emergent or otherwise physically unacceptable, the associated theory of the intension of 'zombie' is less ontologically parsimonious (however exactly such parsimony is construed) than one on which the extension of 'zombie' is null at the relevant worlds. The former theory is also less systematic than the latter, so far as accommodating the possibility that conscious states can have physical effects is concerned, for it is not perfectly clear how nonphysical goings-on can interact with physical goings-on (as was originally pressed by Princess Elizabeth of Bohemia against Descartes; see Kim 2015 for recent discussion); and though my own view is that this can be made sense of via the notion of a fundamental mental interaction, in any case such an account requires further ontology and ideology. One might sidestep this concern by coupling a theory on which 'zombie' has a non-null extension with the supposition that conscious states are epiphenomenal,

as Chalmers sometimes seems to suggest. But (and as is reflected in the dearth of 'live' epiphenomenalist accounts of mentality) such a denial is highly implausible, given our experience, and moreover unsystematic, given that no other empirical goings-on are taken to be epiphenomenal; hence this theory too suffers by way of comparison with one taking 'zombie' to have a null primary intension.

The ultimate abductive comparison of theories of the intension(s) of 'zombie' will involve not just these, but many other considerations. Even without entering further into this issue, however, it is clear that an abductive rather than (merely) conceiving-based approach to the question at hand is very far from indicating that 'zombie' has a non-null extension at some world, as is required on the E2D strategy if zombies are to be possible. If anything, the all-things-considered weight seems likely to be on the other side. To be sure, a positive empirical result, indicating the presence of a novel fundamental interaction at the level of mentality, might trump all these other considerations; but at present, we are lacking such evidence. Hence even granting the independent desirability of the E2D framework as well as the conceivability of zombies, nothing yet follows about whether zombies are metaphysically possible.

Chalmers's conceivability argument, like the knowledge argument(s), thus fails to establish that consciousness and its associated subjective and qualitative aspects are actually physically unacceptable, much less actually Strongly emergent.

7.2 Is consciousness Weakly emergent?

I turn now to considering whether any conscious mental features are plausibly seen as Weakly emergent from the lower-level physical states upon which they depend.

As prefigured, my strategy for answering this question will proceed by motivating the claim that qualitative conscious features—e.g., states of conscious awareness of colors or pains—are reasonably taken to be typically determinable rather than (maximally) determinate, such that (on the assumption that conscious states are not actually Strongly emergent) these determinable conscious states are plausibly seen as actually realized by determinate lower-level physical states, in accord with a determinable-based account of Weak emergence. As we'll see, this application of a determinable-based account faces certain challenges, which I will argue (following Wilson 2009) can be addressed.

7.2.1 Determinable perceptions

Among the most salient conscious mental states are those which we might call 'qualitative', in constitutively involving experience of features such as colors, tastes,

textures, pains, and the like. There are (at least) two motivations for taking the qualitative aspects of conscious experiences to be determinable rather than maximally determinate, rendering the states themselves appropriately seen as determinable rather than maximally determinate.

The first motivation appeals to Sorites phenomena, as indicating that we fail to perceive fully determinate instances of many properties, including colors, tones, and textures. A Sorites sequence consists in a sequence of pairwise indiscriminable instances of such features. An application of the notorious Sorites paradox consists in first registering the plausibility of a principle according to which, if one starts with an instance that is (judged to be), e.g., clearly red, and given that the next patch is indiscriminable from the first, the next patch must also be (judged to be) red. One then iteratively implements this reasoning, starting with the first (clearly red) patch, such that the second, indiscriminable patch must also be (judged to be) red, and similarly for the second and third patches, and so on, in what is sometimes called a 'forced Sorites march'—until one arrives, contradictorily, at a patch that is clearly not red (e.g., a patch that is clearly orange), but which must be (judged to be) red. As Fales (1990) notes, that the qualitative aspects of perceptual experiences are to some extent determinable rather than maximally determinate "is a conclusion which seems forced upon us by the fact that each member of a series of colors, etc., may be perceptually indistinguishable from its immediate neighbors but easily distinguishable from more distant members of the series" (172).

The second motivation reflects that our perception of macro-entities and their features typically fails to register micro-determinate details. As I observed in Ch. 1 (§1.1.2), the macro-entities of our acquaintance do not perceptually appear to us as massively complex, constantly changing, configurations of micro-phenomena. The same is true of the qualitative features of such entities. For example, we do not experience the shapes of macro-entities in fully microscopically determinate detail; rather we experience these shapes as to some extent determinable.

Two further points about these motivations for taking perceptions to be typically determinable are worth noting. First, the motivations here do not depend on the claim that the objects or features perceived are objectively indeterminate. My own view, as per the previous chapter, is that some features of ordinary objects (including but likely not limited to their boundaries) are to some extent metaphysically indeterminate, which when combined with a determinable-based view of metaphysical indeterminacy entails that the features (hence the objects) are in fact indeterminate in the relevant respect(s). But even if the objects and features perceived are themselves completely determinate in all (or the relevant) respects, it would remain that qualitative conscious states involving perceptual experience of such objects would still be less than maximally determinate. As I note in my (2012), in discussing how perception provides motivation for thinking that determinables exist:

> Perhaps the (instances of) properties perceived are really maximally determinate, and only perceptual features or modes of presentation are determinable; but features of perceptual experience are also aspects of reality, so the larger point remains. (5)

It seems reasonable to suppose, then, that at least some qualitative conscious states are less than maximally determinate, owing to these states' having qualitative aspects that are less than maximally determinate.

Second, the phenomena at issue suggest that there is no hope here of providing a deflationary—anti-realist or reductionist—treatment of the determinable qualitative conscious states at issue. That we have qualitative conscious states (qualitative experience) isn't, in my view, up for grabs—it's as epistemically foundational in our experience as any aspect of reality, or more so. Hence any argument for eliminating such states would be less compelling than the sheer fact of our experience.[16] Nor is it plausible to suppose that these determinable states might be given some or other reductive treatment in terms of maximal determinates. There are a number of routes towards the rejection of any such reduction, including that in Ch. 2, §2.3.1, according to which determinables and associated determinates satisfy the *Proper Subset of Powers Condition* (see Wilson 2012 for other arguments for irreducibility). For present purposes it suffices to observe that any reductive treatment of a determinable state which proceeds by identifying a determinable type with a disjunction (or some other 'lightweight' combination) of determinate types entails that every token of a determinable type is identical with a token of a determinate type—and indeed, a token of a maximally determinate type. But the phenomena associated with qualitative conscious states undercut the hope of any such reduction, since as previously, our token qualitative conscious experiences are not maximally determinate. It is thus reasonable to assume that at least some conscious experiences, understood as having constitutive qualitative aspects, must be characterized in irreducibly determinable terms.

Now, as previously, one account of realization of the sort satisfying the conditions in the schema for Weak emergence is a determinable-based account, according to which it suffices for the realization of a feature that the feature be a determinable of lower-level physical determinates. So, if the determinable qualitative conscious states at issue can be seen as having lower-level physical determinates, we will be in position to conclude that such conscious features are Weakly emergent.

[16] This is a common, and to my mind compelling, reason to reject the sort of eliminativism about qualitative mental states endorsed by the Churchlands (in, e.g., Patricia Churchland 1986 and Paul Churchland 1981).

7.2.2 The objections from mental multiple realizability and mental superdeterminates

Ehring (1996), Funkhouser (2006), and Walter (2006) argue, however, that it does not make sense to see lower-level physical states as determinates of determinable conscious mental states.[17] The common line of argument in these discussions is schematically as follows:

1. The determinable/determinate relation has feature *F*.
2. The relation between conscious states and lower-level physical states does not have *F*.
3. Therefore, the relation between conscious states and lower-level physical states is not the determinable/determinate relation.
4. Conscious states are not realized, in determinable-based fashion, in lower-level physical states.

There are two instantiations of the above argument schema relevant to the question of whether qualitative conscious states might have lower-level physical determinates.

The first, associated with 'the argument from mental multiple realizability', appeals to the feature of the determinable/determinate relation according to which determinates of a determinable differ 'in respect of' their determinable. This feature reflects that the distinctive form of specification whereby a determinate is more specific than a given determinable is supposed to contrast with the genus/species relation and, more generally, the conjunct/conjunction relation (see Wilson 2017 for historical discussion). The latter specification relations are compatible with the increase in specificity's involving the conjunctive addition of some independent property, as when a genus is conjoined with some differentium to form a species, or when a state of affairs is made more specific by being conjoined with another state of affairs. The specification at issue in paradigm cases of the determinable/determinate relation is not properly understood in such conjunctive terms. For example, *scarlet* is not appropriately analyzed as a conjunctive combination of *red* and some other property; rather, determinates are in some sense more intimately specific 'in respect of' their determinables. One way in which this more intimate variety of specification gets elucidated is in terms of the determinable's having certain 'determination dimensions', along which the

[17] Worley (1997) and Funkhouser (2006) also argue that certain mental states are not appropriately seen as determinables of lower-level physical determinates; but since these discussions focus on beliefs, they are not directly relevant to the present question of whether qualitative conscious states can stand in a determinable/determinate relation to lower-level physical states. See Wilson 2017 for discussion of Worley's and Funkhouser's concerns.

determinable can be rendered more determinate (as per Funkhouser 2006). For example, it is common to suppose that the determination dimensions of *color* are hue, saturation, and brightness, with different determinates of *color* differing in respect of color by differing in respect of how they are specified along one or more of these determination dimensions.

Now, according to the argument from mental multiple realizability, that conscious states are typically multiply realizable rules out that lower-level physical states differ in respect of any such conscious state, hence rules out taking conscious states to be determinables of lower-level physical determinates:

> [T]he physical realizers of the mental will not differ mentally at all, as they should if they are determinates of the requisite mental states. (Ehring 1996, 474)

> Mental properties are said to be multiply realizable precisely because distinct physical realizers can be exactly the same with respect to the mental property they realize. (Walter 2006, 219)

In particular, such multiple realizability is supposed to be characteristic of qualitative conscious states. Note that the problem being raised here is not that we can't make sense of determinates of a given qualitative conscious state as being exactly the same with respect to the mental property they realize—there's no problem with this (perhaps, e.g., the multiple determinates are exactly similar in sharing the determinable or its powers, or in their essences's containing the same determinable essence, as Yablo suggests in his 1992). Rather, the concern is that there is no clear way to see qualitative conscious states as having determination dimensions that can be further specified by their multiple lower-level physical realizers. Hence, for example, if the determination dimensions of perceived colors are hue, saturation, and brightness, there is no clear way of understanding how perceived values of these features could be further specified by lower-level physical realizers: it's not as if the lower-level physical realizers are capable of realizing more precise color perception! And similarly for other qualitative conscious states, such as pain.

The second instantiation of the above argument schema, associated with 'the argument from mental superdeterminates', relies on the feature of the determinable/determinate relation according to which some determinables admit of maximal specification. In such cases, there is a maximal determinate or 'superdeterminate' of the determinable. According to the argument from mental superdeterminates, some qualitative conscious states are superdeterminate:

> [S]uppose that *M* is a fully determinate type of mental state. For example, make *M* a precise state of searing pain such that there is no room for further specification of this mental state *qua* pain state. [...] Suppose that *M* and "being in pain" have a

> physical [realizer], *P*. Suppose that *P* is a determinate of *M* and "being in pain". [If *M* is realized by *P* in determinable-based fashion] *M* cannot be a fully determinate pain state. This is so because there are further determinates of that pain state [...] in the form of *P*. But in fact, *M* is a fully determinate pain state by hypothesis. Thus the physical [realizers] of mental properties are not determinates of that which they realize since if that were true, *M* would not be a fully determinate pain property. (Ehring 1996, 473)

Here the concern is clear—namely, that taking qualitative conscious features to be realized in determinable-based fashion by lower-level physical features falsely implies that certain qualitative mental superdeterminates can be further determined.

7.2.3 Responses to the objections

These concerns are important, but as I will now argue, they can be addressed, given a proper understanding of the determinable/determinate relation.

A powers-based account of determination

Determinables (of whatever ontological category) are less specific than their determinates. In Wilson 1999, I argued that this increase in specificity reflects a proper subset relation between the sets of powers of the types of features involved, as follows:

> *Powers-based Determination* (first pass): Feature *P* is a determinate of feature *Q* iff the set of powers associated with *Q* is a proper subset of the set associated with *P*.

Here the idea is that a determinate is more specific than its determinable in being associated with a more specific set of powers. Hence it is, for instance that in virtue of being scarlet, a patch can do more (say get Alice the picky pigeon to peck at it) than the patch can do simply in virtue of being red.

This first-pass proposal has the virtue of ensuring that there is a contrast between the determinable/determinate relation and the disjunction/disjunct relation (another specification relation with which determination is traditionally taken to contrast), for as previously discussed in Ch. 2 (§3.2.2), disjunctions are associated with more powers than associated disjuncts. As it stands, however, the first-pass proposal does not ensure the contrast between the determinable/determinate relation and the genus/species and conjunct/conjunction relations. Again, it is crucial to the determinable/determinate relation that the determinate cannot be understood or analyzed as a conjunction of the determinable and some other property.

We may preserve this intended contrast on a powers-based approach (improving on the first-pass proposal) by stipulating that the powers in the complement of the sets associated with a determinable and any of its determinates, respectively, are not associated with a set that is associated with any property, as per:

> *Powers-based Determination* (second pass): feature *P* is a determinate of feature *Q* iff *Q* is associated with a proper subset of the powers associated with *P*, and the set of powers had by *P* but not by *Q* is not associated with any property.

The second pass account characterizes a specification relation that appropriately contrasts with both the disjunction/disjunct and conjunct/conjunction relations.

Addressing the objection from multiple mental realizability

Recall that the deeper concern at issue in the objection from multiple mental realizability reflects, first, a conception of the intimate ('in respect of') form of specification associated with the determinable/determinate relation as involving increased specification along one or more determination dimensions of the determinable; and second, the supposition that diverse physical realizers of a qualitative conscious state cannot differ from each other along mental determination dimensions of the conscious states at issue, contra the 'in respect of' feature of the determinable/determinate relation.

My response proceeds by providing empirical reason to think that what determination dimensions are associated with a given determinable conscious state (perceived color, pain) is science-relative, in such a way that we can make sense of such states as having purely 'psychological' determination dimensions relative to certain sciences (e.g., 'normal' color science or some branch of psychology), and as having physical determination dimensions relative to other sciences (e.g., metameric color science or pharmacology). Such relativization makes in-principle room for multiple physical realizers of a qualitative conscious state to have explicitly physical as well as psychological determination dimensions. I then use *Powers-based Determination* to fill in how this might be, from a metaphysical point of view.

Take perceived color, for example. Hue, saturation, and brightness suffice to characterize perceived colors in normal light conditions, as experienced by normally sighted creatures more or less like us. Interestingly, however, things that appear to be the same color under normal light conditions may appear to be different colors under different light conditions. The explanation for this phenomenon—'metamerism'—has to do with broadly physical features of the objects and light at issue.[18] Most notably, what color we perceive an object to be will be a function of the spectral power distribution (SPD) of the light hitting the retina, specifying the power of the light at each wavelength in the visible

[18] See, e.g., Wandell 1993.

spectrum; this SPD is itself a function of the SPD of the light incident on the object's surface, and the surface reflectance properties of the object. Different SPDs of light hitting the retina may give rise to the same 'tristimulus values' (effectively: hue, saturation, and brightness); hence it is that samples that appear the same in normal light conditions may appear different in other conditions, and that samples that appear different in normal light conditions may appear the same in other light conditions. Metamers are color appearance properties that are individuated, in part, by the relevant broadly physical features—let's assume these are the retinal SPDs—needed to accommodate the phenomenon of metamerism, such that color appearance properties not distinguished by hue, saturation, and brightness are distinguished by the relevant broadly physical features.

As I argue in detail in Wilson 2009, metamers are reasonably taken to be colors—just colors seen (no pun intended) through a lens of a finer and partly physical grain. Here I'll just mention two considerations in favour of this claim. First, metamers are part of the broader field of color science, and they are characterized as colors in that science. In particular, color science is not concerned only with colors as individuated by hue, saturation, and brightness.[19] On the contrary, considerable color research is aimed at understanding colors as individuated by broadly physical features such as retinal SPDs, as relevant to digital photography, screen displays, car interiors, etc.[20] Second, the role that retinal SPDs play in this research appears to be compatible with taking colors to be themselves characterized by the relevant broadly physical features. After all, as above, colors are understood as perceptual properties; and retinal SPDs are clearly part of the process of color perception—in particular, retinal SPDs are input into the color-sensitive cones, which then output the tristimulus values. Whether or not the input/output function here is causal or rather 'filter-like', in any case there seems to be no in-principle barrier to characterizing (specific kinds of) colors in terms of the broader process of visual perception—especially since the broader process and associated features are required to *fully* characterize color appearances (in particular, in light of metamerism).

Similar remarks apply to other qualitative conscious states, such as pain. For example, Funkhouser (2006) suggests that states of pain have mental determination dimensions along lines of feel and intensity; but it seems reasonable to suppose that pains that are exactly similar with respect to these psychological determination dimensions might be furthermore physically specified, as sciences such as pharmacology take for granted.

I draw two morals from these sorts of case studies. First is that qualitative conscious states, such as states of perceiving colors or experiencing pains, may have

[19] See, e.g., Wyszecki and Styles 1982.

[20] See, e.g., Judd and Wyszecki 1975.

physical as well as psychological determination dimensions. Second is that determination dimensions may be science-relative: different sciences may treat a given determinable feature as having different determination dimensions, effectively characterizing the determinable feature in more or less fulsome ways. Relative to normal appearance color science, *color* has determination dimensions of hue, saturation, and brightness; relative to metameric color science, *color* has further determination dimensions. Note that there's nothing mysterious about taking determination dimensions to be science-relative, from a physicalist perspective of the sort we are presently considering. That different sciences may treat the same determinable as having different determination dimensions simply reflects that different sciences and their associated laws may treat certain phenomena at different levels of metaphysical grain.

The observation that contemporary science suggests that conscious mental states may have physical as well as psychological determination dimensions provides a wedge for responding to the objection from mental multiple realizability. Still, the question remains: can we make metaphysical sense of this, compatible with qualitative conscious states' being determinables of lower-level physical determinates?

Here's where *Powers-based Determination* comes in.[21] The basic idea is that the phenomenon of a given determinable's being associated with increasingly fine-grained determinable dimensions, relative to a given science, can be straightforwardly understood in terms of nonconjunctive specification of a determinable's powers. Relative to purely psychological determination dimensions, reflecting sensitivity to the set of powers associated with a multiply realized determinable, determinates of the determinable may be exactly alike. Relative to a finer-grained set of determination dimensions, reflecting sensitivity to powers additionally possessed by the physical realizers of the determinable, determinates of the determinable will not be exactly alike. On this approach, what it is for determinates to determine a determinable 'in respect of' a determinable in the sense relevant to ensuring the distinctive form of specification at issue in the determinable/determinate relation is cashed in terms of the determinates' sharing the powers of the determinable (associated with the psychological determination dimensions), but differing with respect to powers going beyond these (associated with the more fine-grained determination dimensions encoding physical as well as psychological distinctions), where the additional powers associated with any given realizer do not themselves form a property. All this is compatible with the multiple realizers' being exactly similar with respect to the determinable's psychological

[21] As I discuss in my (2009), there are likely other approaches to the determinable/determinate relation (as per, e.g., the proposal set out in Funkhouser 2006) that can make sense of the science-relativity of determination dimensions.

determination dimensions in virtue of sharing, at both the type and token levels, the powers had by the determinable, while differing in respect of other powers—and correspondingly, with qualitative conscious mental states being Weakly emergent, as per a determinable-based account of realization.

Accommodating mental superdetermination

Recall that the concern at issue in the objection from mental superdeterminates is that taking qualitative conscious features to be realized in determinable-based fashion by lower-level physical features is incompatible with the intuitive possibility of there being qualitative mental superdeterminates, since implying, falsely, that these could be further determined.

My response again starts with the observation that different sciences may treat a single determinable as having different determination dimensions, such that mental features may be superdeterminate relative to a purely psychological science, while being further determined relative to a lower-level science. What is superdeterminate relative to one science may not be superdeterminate relative to another.

This much provides a wedge for responding to the objection from mental superdeterminates. Still, the question remains: can we make metaphysical sense of this, compatible with qualitative conscious features being determinables of lower-level physical determinates?

Again, *Powers-based Determination* provides a comprehensible metaphysical basis for accommodating the phenomenon at issue. What counts as a superdeterminate depends on which determination dimensions are at issue. Relative to one set of determination dimensions, reflecting sensitivity to powers associated with the determinable set, a given qualitative conscious state might be characterized as a superdeterminate. Relative to a finer-grained set of determination dimensions (reflecting sensitivity to powers in relevant supersets of the determinable set) that same feature might not be appropriately characterized as a superdeterminate.

All this is compatible both with certain qualitative conscious features being superdeterminate with respect to certain psychological determination dimensions, but being further determinable with respect to further, partly physical determination dimensions—and so with qualitative conscious mental features being Weakly emergent, as per a determinable-based account of realization.

7.3 Concluding remarks

Let's sum up the results of this chapter. Consciousness—more specifically, the sort of qualitative or 'what it's like' mental features characteristic of conscious

experience—has been frequently offered up as the best candidate for an actual Strongly emergent phenomenon. But, I've argued, neither the knowledge argument(s) nor the conceivability argument establishes that qualitative conscious features are actually physically unacceptable, and so they provide no basis for moreover taking such features to be actually Strongly emergent. While the Strong emergence of qualitative consciousness remains an open empirical possibility, the presently best case for the metaphysical emergence of such consciousness is one according to which, due to the irreducibly determinable nature of qualitative features of conscious experience, such features are Weakly emergent, by lights of a determinable-based account. As I've argued, such an understanding of qualitative conscious features can be defended against various objections, given a proper appreciation of the science-relativity of determination dimensions and a proper understanding of the determinable/determinate relation, as per *Powers-based Determination*. I conclude that consciousness—more specifically, qualitative conscious experience—is at least Weakly emergent.

8
Free will

Free will (or free agency), if such there be, involves the ability to mentally choose an outcome (an intention to φ, or a φ-ing), where the outcome is 'free' in being, in some substantive sense, up to the agent of the choice. Free will has often been taken to be core to what it is to be a person, of either human or other variety, in part (though not exclusively) because such agency seems to be prerequisite for persons' being autonomous in the way seemingly crucially relevant to achieving certain moral, aesthetic, and other goods. In this chapter, I consider reasons for thinking that free will, as it actually exists, might be either Weakly or Strongly emergent.

I start by drawing on Bernstein and Wilson 2016 in order to set up a useful framework for investigating into whether free will is metaphysically emergent. Recall, to start, that the schemas for Weak and Strong emergence were initially motivated as associated with two specific responses to the problem of higher-level causation; namely, nonreductive physicalism and Strong emergentism. A common focus of this problem concerns the status of mental features; but in the usual case the mental features at issue are qualitative or intentional features, for which free choice is supposed not to be at issue. More generally, debates over the status of free will have tended to proceed in relative independence from debates over the status of mental features whose governance by natural law is taken for granted. As Bernstein and I argue, however, the problematics underlying the free will and the mental causation debates are appropriately seen as special cases of a more general problem, concerning whether and how mental features of a given type may be efficacious, *qua* the types of feature they are (qualitative, intentional, freely deliberative), given their apparent causal irrelevance—i.e., apparent failure of distinctive efficacy—for effects of the type in question. (The literature I will be discussing typically focuses more specifically on mental *events* of a given type, so I will present the parallel in these terms.)

That the free will and mental causation debates can be seen as special cases of a more general problem serves to suggest certain parallels between positions in the respective debates, which parallels are useful for purposes of assessing whether free will is either Weakly or Strongly emergent. In particular, as Bernstein and I argue, a representative range of compatibilist (or 'soft determinist') accounts of

Metaphysical Emergence. Jessica M. Wilson, Oxford University Press (2021). © Jessica M. Wilson.
DOI: 10.1093/oso/9780198823742.003.0008

free will[1] implement a strategy that is structurally similar to the 'proper subset' strategy that, I have argued, is core and crucial to nonreductive physicalist accounts of realization, and more generally, to Weak emergence. Here I present and extend this result, arguing that the compatibilist's proper subset strategy is reasonably taken to indicate that compatibilist free will, were it to exist, would be Weakly emergent. I then argue that a representative range of libertarian treatments of free will are appropriately seen as committed to such agency's involving a fundamentally novel power, such that libertarian free will, were it to exist, would be Strongly emergent. Against this backdrop, I go on, in the final section, to consider the prospects of there being metaphysically emergent free will, suggesting that free will of the compatibilist, Weakly emergent variety is common, and providing a new argument for there moreover actually being free will of the libertarian, Strongly emergent variety.

8.1 The generalized problem of mental quausation

In our (2016), Bernstein and I argue that (certain understandings of) the problems of free will and of mental causation can be seen as special cases of a more general problem, concerning whether and how mental events of a given type may be efficacious, *qua* the types of feature they are—qualitative, intentional, freely deliberative, and so on—given their apparent causal irrelevance for effects of the type in question.

In making this connection, we generalize what Horgan (1989) calls "the problem of mental quausation".[2] As Horgan presents it, this problem is a refinement of the problem of mental causation (a special case of the problem of higher-level causation presented in Ch. 2, §2.1.1). The problem of mental causation is sometimes pitched as the problem of how a (real, distinct, cotemporally materially dependent) qualitative or intentional mental event M might be efficacious *at all*, given that any effect E it might be seen as causing is, by *Physical Causal Closure*, already caused by the physical event P upon which M depends. But as Horgan notes, there is a quick route to gaining M's efficacy—namely, via identifying M with P, as per the usual reductive physicalist response—that leaves open what is arguably the deeper question underlying the original problem:

[1] In general, compatibilists maintain, as Lewis (1981) puts it, that "soft determinism may be true" (113). Beyond this commitment, individual compatibilists may aim just to carry out the 'negative' project of undermining arguments for incompatibility, or in addition to provide a positive conception of the compatibility at issue (see McKenna and Coates 2008 for discussion). Here by 'compatibilism' I mean 'positive compatibilism'.

[2] Though Horgan's (and our) focus is on mental goings-on, the more general problem of 'quausation' can be seen as applying to any seemingly higher-level goings-on.

> Even if individual mental events and states are causally efficacious, are they efficacious *qua* mental? I.e., do the mental types (properties) tokened by mental events and states have the kind of relevance to individual causal transactions which allows these properties to figure in genuine causal explanations? (47)

What is needed to adequately vindicate the efficacy of the mental, Horgan suggests, is that mental events be shown to be distinctively efficacious: efficacious *qua* mental—and more specifically, '*qua F*', where "*F* is schematic for a specific type of mental event" (50). As such, it is mental 'quausation', not mental causation per se, that is most deeply challenged by the possible truth of *Physical Causal Closure*. To wit: how can a mental event *M* be efficacious vis-à-vis an effect *E*, in virtue of being qualitative or intentional, given that *E* is causally determined by physical events and associated laws in ways that seem to preclude *M*'s being causally relevant in either of these (qualitative or intentional) respects?

Though Horgan's discussion targets mental events for which freedom is not at issue, the deeper concern about whether and how mental events can be seen as causally relevant—that is, efficacious in virtue of, or *qua*, the distinctive mental types they are—lies also at the core of what is sometimes called 'the consequence argument', 'the problem of free will and determinism', or just 'the problem of free will'. This problem highlights a seeming tension between an intuitive conception of free will as involving the ability to freely choose an outcome φ, and the broadly scientific thesis of *Determinism*, according to which every event is a consequence of the laws of nature and the state of the world at any time.[3] *Determinism* admits of different interpretations;[4] what is at issue here is a reading involving causal determination of present or future events by prior states (broadly construed), as follows:[5]

[3] One might wonder whether *Determinism* can be ruled out of court, on grounds that quantum mechanics, which is likely true, is indeterministic. This would be premature, however, both because there are deterministic interpretations of quantum mechanics (see, e.g., Bohm 1952), and because the present incompatibility of quantum mechanics and general relativity suggests that we are not quite yet in position to infer to the likely truth of quantum mechanics. (See O'Connor 2005, 4 for similar observations.) Of course, even if *Determinism* is actually false, one might be interested in understanding what bearing the truth of *Determinism* would have on free will. More importantly, the problem of free will doesn't crucially depend on the assumption of *Determinism*, since as I'll discuss in what follows, concerns about the causal relevance of free will also arise if the outcomes of seemingly free choices are the product of indeterministic (e.g., quantum) laws. The focus on *Determinism* is traditional, and convenient for setting up the parallel with the problem of mental quausation, but is ultimately dispensable.

[4] Steward (2015) distinguishes between interpretations based on entailment between propositions at one time by propositions at another time, and interpretations based on causation between events.

[5] Our focus here thus excludes compatibilists whose rejection of causation or causal production between events—Leibniz, and perhaps also Hume—would entail their rejection of a causal reading of *Determinism*.

> *Causal Determinism*: With the exception of any first events there might be, every event is a causal consequence of the laws of nature and the state of the world at any prior time.

The possible truth of *Causal Determinism* leads to a question: If every event *E* (e.g., an intention to φ, or a φ-ing) purportedly caused by a mental event of free choosing *M* is, by *Causal Determinism*, a causal consequence of the laws of nature and prior states, what causal role is left for *M* to play, vis-à-vis *E*? The answer, according to (a causal reading of) the consequence argument, is that no role is left. As van Inwagen (1983) puts it:

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.[6] (16)

The question and prima facie negative answer constitute what van Inwagen calls "the problem of free will and determinism", and what Bernstein and I call, for short, 'the problem of free will'.

In the problem of free will, it is not the mere efficacy of deliberative mental events (as, perhaps, the most proximate causes of intentions or actions in a causally determined chain of events) that is at issue; indeed, that events of choosing are efficacious is typically taken for granted in the free will debates. Rather, what is at issue is whether events of choosing can be efficacious *qua* free—again, in eventuating in outcomes that are, in some substantive sense, up to the agent—under circumstances where *Causal Determinism* is presumed to be true. How could an event of mental choosing *M* be efficacious vis-à-vis an event *E*, in virtue of being *free*, given that *E* was causally determined by laws of nature and events antecedent to *M* in ways that seem to preclude *M*'s being causally relevant in this respect?

This parallel suggests that the problem posed by *Physical Causal Closure* for qualitative and intentional mental events, and the problem posed by *Causal Determinism* for events of seemingly free choosing, are each instances of a suitably general problem of mental quausation:

> How can a mental event *M* of a given type be efficacious *vis*-à-*vis* an event *E*, in virtue of being the type of mental event it is, given that there is reason to think that events of *M*'s type are causally irrelevant to the production of events of *E*'s type? (Bernstein and Wilson 2016, 314)

[6] Van Inwagen's remarks are directed at an entailment-based reading of *Determinism*, but they apply, mutatis mutandis, to a causal reading (simply insert 'causal' before 'consequences').

Note that in drawing this parallel, Bernstein and I are not suggesting that there is any interesting parallel between the problem of free will and the *unrefined* problem of mental (more generally, higher-level) causation; nor are we suggesting that the problem of free will is an instance of the problem of mental quausation as applied to qualitative or intentional mental events; after all, there is no clear tension between *Causal Determinism* and mental quausation involving mental causes for which freedom is not at issue. Our point is simply that both the problem of free will and the refined problem of mental causation (again, typically directed at qualitative and intentional mental events for which free will is not at issue) are specific cases of a suitably general problem of mental quausation, whereby the causal relevance (distinctive efficacy) of a mental event of a distinctive type—whether intentional, qualitative, or freely deliberative—is called into question by the holding of certain theses (*Physical Causal Closure*, *Causal Determinism*) which are to some extent open possibilities for being true.[7]

The traditional responses to the problem of free will may be categorized by reference to the following free will conditional:

> If all events are subsumed by deterministic natural laws, then free mental quausation—the causation of events (e.g., intentions to φ, φ-ings) by mental choosings *qua* free—does not exist.

Hard determinists take the conditional to be true, on grounds that both antecedent and consequent are true; libertarians take the conditional to be (vacuously) true, on grounds that the antecedent is false; soft determinists or 'positive' compatibilists (henceforth, just 'compatibilists') reject the conditional, on (the usual) grounds that the antecedent is true but the consequent is false. Similarly, the traditional responses to the refined problem of mental causation can be categorized by reference to the following mental causation conditional:

> If all physical events are subsumed by physical laws, then qualitative and intentional mental quausation—the causation of physical events (e.g., bodily movements) by mental events *qua* qualitative or intentional—does not exist.

Eliminativist physicalists and epiphenomenalists take the conditional to be true, on grounds that both the antecedent and consequent are true; Strong emergentists and substance dualists take the conditional to be (vacuously) true, on grounds that the antecedent is false; reductive physicalists and nonreductive physicalists/Weak

[7] Again, though it is convenient to present the parallel in terms of the problem for free will posed by the assumption of *Determinism*, concerns can also be raised (as I'll discuss down the line) for the causal relevance of free will on the assumption that free choices are the product of indeterministic laws.

emergentists reject the conditional, on grounds that the antecedent is true but the consequent is false.

Since the problem of free will and the refined problem of mental causation may each be seen as special cases of a more general problem, we might expect there to be parallels between the primary responses to each problem, corresponding to parallels in the stances taken towards the components of the corresponding conditional. As I will argue in the following two sections, this is indeed the case for compatibilism and libertarianism, respectively.[8]

8.2 Compatibilism and Weak emergence

According to compatibilism, the free will conditional is false, since the antecedent (*Causal Determinism*) is true (or might be true), but the consequent (denying the existence of free will) is false. Here two positions—reductive and nonreductive physicalism—take the same stance as regards the refined mental causation conditional. As previously discussed, however, reductive versions of physicalism, which identify the mental events at issue with physical events, face immediate difficulties in making sense of how mental events can be efficacious *qua* the qualitative or intentional types they are, with the usual strategies (as per the deflationary accounts discussed in Ch. 3, §3.1, §3.2) being to offer pragmatic or purely epistemic accounts of the desired 'higher-level' efficacy. Given this, and since the deeper concern at issue in the problem of free will is with whether there is a genuinely metaphysical basis for free mental quausation, we can cut to the chase of considering whether there is an interesting parallel between compatibilism and nonreductive physicalism/Weak emergentism.

8.2.1 The compatibilist's proper subset strategy

At this point, the reader is well familiar with the characteristic strategy of the nonreductive physicalist's response to the problem of higher-level quausation, as encoded in the schematic conditions in *Weak Emergence*:

> *Weak Emergence*: What it is for token feature S to be Weakly metaphysically emergent from token feature P on a given occasion is for it to be the case, on that occasion, (i) that S cotemporally materially depends on P, and (ii) that S has a non-empty proper subset of the token powers had by P.

[8] There is also a parallel between hard determinism and eliminativism; see Bernstein and Wilson 2016 for discussion.

As setup for motivating the structural parallel with compatibilism, it is worth recalling how satisfaction of these conditions—most crucially, satisfaction of the *Proper Subset of Powers Condition*—operates to secure, not just the reality and efficacy of the higher-level goings-on, but also their causal relevance/distinctive efficacy. Again, the core idea is that there are two ways for a higher-level feature *S* to be distinctively efficacious with respect to its base feature *P*. One way is for *S* to be associated with a *new* power—a power that *P* doesn't have, or doesn't have in the same (direct) way as *S*; this is the form of distinctive efficacy at issue in Strong emergence. A second way reflects that *S* is associated with a *distinctive power profile* consisting of a proper subset of the powers associated with *P*; this is the form of distinctive efficacy at issue in Weak emergence.

Recall also that there are (at least) two strategies which might be used to motivate taking the having of a distinctive power profile conforming to Weak emergence to provide a basis for *S*'s being efficacious *qua* the type of higher-level feature it is: first, by observing that such a power profile provides a basis for difference-making or 'proportionality' considerations (where the effect would still have occurred and been caused by *S* even had *S* been differently realized); second, by observing that such a power profile may be associated with systems of laws or causal joints that are comparatively abstract, in being insensitive to certain details needed for the lower-level physical laws governing *P* to operate. Applied to the qualitative or intentional mental events associated with the problem of mental quausation: even if every token power of a given such mental event *M*, on a given occasion, is identical with a token power of its physical realizer *P* on that occasion, *M* can be distinctively efficacious—causally efficacious *qua* qualitative or intentional—in virtue of *M*'s distinctive power profile's tracking either difference-making considerations, or the comparatively abstract system(s) of laws or causal joints associated with (non-agential) psychology.

Interestingly, as Bernstein and I argue, a representative range of seemingly diverse compatibilist accounts implement a structurally similar 'proper subset' strategy for responding to the problem of free will. To prefigure: in the case of compatibilist free will, the operative strategy involves characterizing events of seemingly free choosing as associated with only a proper subset of the causal determinants of the outcome (effect) at issue, in such a way as to provide a principled basis for the claim that the choosing is efficacious *qua* free. Hawthorne and Pettit's (1996) taxonomy of compatibilist strategies serves as a useful basis of operations for establishing this result. They start by noting:

> All compatibilists agree that every choice has antecedents and [...] that this fact puts freedom of choice in doubt. How can a choice be made freely if it is the product of independent antecedents? The response they make is that some possible antecedents are better than others from the point of view of free

> choice and that a choice is free to the extent that its antecedents, or at least its relevant antecedents, satisfy the inherently vague condition of leaving it up to the agent. (191)

In schematic form:

> *X* chooses freely to φ if and only if the relevant antecedents of the choice leave the φ-ing up to *X*. (191)

As above, for present purposes a choice to φ (the outcome of an event of choosing) may be either an intention to φ, or a φ-ing. Hawthorne and Pettit identify three main compatibilist accounts of what it is for a choice to be "up to an agent", associated with the notions of freedom as underdetermination, ownership, and responsibility, respectively. The accounts vary to some extent as regards which antecedents are supposed to be relevant to establishing whether the choice was up to the agent, in the intended sense. As Bernstein and I argue, however, each of these accounts plausibly imposes satisfaction of a certain 'proper subset' condition, as key to the associated strategy of response to the problem of free will.

I start by motivating the condition. First, reflecting endorsement of (the open possibility of) *Causal Determinism*, the compatibilist accepts the following condition on the causal antecedents of any outcome of a free mental deliberation:

> *Causal Antecedents Condition*: The total causal antecedents of an event of free choosing *M* completely determine the outcome of *M* (e.g., a choice to φ).

As a first pass, the compatibilist strategy requires that a free mental choosing *M* satisfy the following proper subset condition:

> *Subset of Causal Antecedents Condition* (first pass): The relevant causal antecedents $\{C\}$ of a free mental choosing *M* are (i) a non-empty proper subset of the total causal antecedents of *M*, which (ii) satisfy the condition of leaving the outcome of *M* up to the agent.

Moreover, as will shortly become clear, if the compatibilist's strategy of identifying the relevant causal antecedents of *M* is to make sense of the idea that these antecedents leave the choice up to the agent, then the relevant antecedents must more specifically satisfy the following (final pass) proper subset condition:

> *Proper Subset of Causal Antecedents Condition*: A free mental choosing *M* resulting in a choice to φ satisfies the following: (i) *M* has relevant causal antecedents $\{C\}$ which are a non-empty proper subset of the total causal antecedents of *M*, and

> (ii) it is possible that a choosing M' of the same type as M occur, having relevant antecedents $\{C'\}$ of the same type as $\{C\}$, but where the *total* antecedents of M' are such as to completely determine the outcome of M'[9] as either a choice not to φ or as the absence of a choice to φ. (Bernstein and Wilson 2016, 325)

After sketching Bernstein's and my arguments that the following three compatibilist accounts aim to satisfy this condition, I will go beyond the previous discussion, to make the case that the compatibilist's *Proper Subset of Causal Antecedents Condition* is not just structurally similar to, but can moreover be plausibly understood as entailing, the *Proper Subset of Powers Condition* operative in the schema for Weak emergence. This will set up for the later discussion of whether compatibilist free will (hence persons characterized as having such agency) is Weakly emergent.

Freedom as underdetermination

On underdetermination accounts, a choice to φ is the result of a free choosing M iff M could have resulted in a choice not to φ. How could this be, given that the choice to φ was determined, as per the *Causal Antecedents Condition* (encoding *Determinism*)? The underdetermination approach proceeds by identifying a subset {C} of the causal antecedents of the choice to φ, relative to which it was left open whether or not M would result in a choice to φ. As Hawthorne and Pettit (1996) put it:

> Taken as a whole, the antecedents of any choice will necessitate that choice under a deterministic picture and compatibilists of this stripe must take the relevant antecedents to be a subset of the totality. But which subset? (193)

The relevant subset of antecedent events will include the choosing event, along with events tracking the standing beliefs and desires of the agent at the time of choosing, and events tracking whether the choosing took place under conditions of physical restraint, threat, etc. (cf. Ayer 1954). From broader perspectives, the relevant antecedents might also include events tracking cultural influences, past trauma, or other psychological, social, psychiatric, neurological, etc., conditions holding of the agent. In general, Hawthorne and Pettit note,

> The line will be that an agent is free to the extent that the antecedents that can or have to be countenanced in that perspective leave the choice underdetermined. [...] To be free, if you like, is to be free relative to that stance. (193–4)

[9] As per the *Causal Antecedents Condition*; the possibility of indeterministic scenarios is put aside here.

Here the relevant antecedent events must be a *proper* subset of the causal antecedents that, as per the *Causal Antecedents Condition*, completely determine the choice, since only if the subset is proper is there any hope that the subset of antecedents will leave that choice underdetermined. Moreover, the assumption that the subset of antecedents leaves the choice underdetermined plausibly entails (indeed, has as its content) that it is possible that a choosing M' of the same type as M occur, having relevant antecedents C' of the same type as M's relevant antecedents C, but where the *total* antecedents of M' are such as to completely determine a different outcome, as either a choice not to φ or the absence of a choice to φ.

This last entailment just is the *Subset of Causal Antecedents Condition*. As such, a freedom as underdetermination form of compatibilism explicitly implements a proper subset strategy, characterizing a mental choosing M as associated with a proper subset of its causal antecedents, then using this association to accommodate M's being causally relevant/distinctively efficacious *qua* free vis-à-vis the ensuing choice.

Freedom as ownership

A second compatibilist approach takes freedom to be a matter of ownership:

> The ownership line takes a choice to be up to an agent to the extent that it is not due to anyone or anything other than the agent themselves; it is a choice that the agent owns, a choice with which the agent identifies, and not something forced upon them. Suppose that the relevant antecedents in the adjudication of free will are taken to be […] beliefs and desires. [Then] an agent φs freely just in case their beliefs and desires combine to lead—or at least lead in 'the right way' (see Davidson 1963) to their φ-ing. (Hawthorne and Pettit 1996, 194)

Underdetermination by the relevant antecedents is not explicitly required here, since an agent could own or identify with completely determined intentions. But, Hawthorne and Pettit argue, if the 'ownership' line is to be viable, it will have to ensure underdetermination by these antecedents (hence, Bernstein and I maintain, satisfaction of the *Subset of Causal Antecedents Condition*).

Why so? To start, note that if, for example, I am brainwashed with beliefs and desires leading to my choice to φ, this intention cannot be seen as the effect of free mental choosing. A well-known response (see Frankfurt 1971) requires that choices result from desires that the agent X desires, at the second order, to have and be moved by. The brainwashing problem will re-arise, however, unless "the action issues from desires that the agent has some measure of second-order control over" (Hawthorne and Pettit 1996, 195). As O'Connor (2005) puts it:

> We can [...] imagine external manipulation consistent with Frankfurt's account of freedom but inconsistent with freedom itself. [...] one might discreetly induce a second-order desire in me to be moved by a first-order desire—a higher-order desire with which I am satisfied—and then let me deliberate as normal. Clearly, this desire should be deemed "external" to me, and the action that flows from it unfree. (12)

These considerations indicate that one needs to ensure, somehow, that the formation of second-order desires is up to the agent; and the natural compatibilist approach will be to restrict attention to a proper subset of the antecedents determining the desire—e.g., those relevant to whether the agent's choosing was constrained by other persons, or by other psychological, social, psychiatric, neurological, etc., conditions. In other words, to accommodate the needed control of second-order desires on a 'freedom as ownership' picture, a proper subset of the antecedents of the choice to φ must be specified, relative to which it was underdetermined that the agent had the second-order desires they had; hence underdetermined that the agent would identify with the first-order beliefs and desires leading to the agent's choice to φ; hence underdetermined that the agent's choosing would result in a choice to φ. Such underdetermination in turn entails (indeed, has as its content) satisfaction of the *Subset of Causal Antecedents Condition*.

More generally, here again a proper subset strategy is implemented, whereby a mental choosing M is associated with a proper subset of its causal antecedents, in service of making room, as per the *Subset of Causal Antecedents Condition*, for M to be causally relevant/distinctively efficacious *qua* free vis-à-vis the ensuing choice.

Freedom as responsibility

Lastly, on the 'freedom as responsibility' approach, a choice to φ is the result of a free choosing M iff the agent of the choosing could be held responsible for the outcome of M. The criteria for an agent's being responsible might advert to prevailing systems of law and morality, or (as per Strawson 1962) to the participant or reactive attitudes characteristic of human interactions. As Hawthorne and Pettit point out, this approach again requires that the choice at issue be underdetermined:

> To hold an agent responsible in certain choices is to think that it is not inevitable either that they get things right or that they get them wrong—either that they do well or that they do ill—and so it is to believe that there is a sense in which they could have done otherwise [...]. (197)

The relevant antecedents in this case would then include those relevant to determining whether the agent was deliberating under conditions where they would, by

the lights of the prevailing system of law, morality or interaction, be held responsible for the outcomes of their choosings. Once more, these might cite events or states tracking whether various physical, psychological, neurophysiological, etc., conditions or constraints were in place antecedent to or concurrent with the choosing.

Here again, the account identifies a relevant proper subset of *M*'s causal antecedents, then requires that, relative to these antecedents, the outcome of *M* could have been different, as a way of making sense of *M*'s being causally relevant/distinctively efficacious *qua* free vis-à-vis the ensuing choice. In other words, a freedom as responsibility account implements a proper subset strategy, encoded in satisfaction of the *Subset of Causal Antecedents Condition*.

8.2.2 Deepening the parallel: a powers-based interpretation of the compatibilist's proper subset condition

Nonreductive physicalist and compatibilist positions thus each respond to the respective problems of mental quausation by characterizing the mental events at issue as associated with a proper subset of the 'causal determinants' of their associated effects. As Bernstein and I present the structural similarity, the determinants are not the same: in the one case, these are powers; in the other, these are causal antecedents. This much establishes a structural similarity in strategies: in each case, associating the mental event *M* with (only) the relevant proper subset of causal determinants provides a basis for showing that *M* is causally relevant/distinctively efficacious vis-à-vis the production of the effect *E* in question, *qua* the type of mental event *M* is.

As I will now argue, however, the parallel is even deeper: the compatibilist strategy can be understood as entailing the holding of a proper subset relation between token powers associated with two complex, cotemporal events, corresponding to, first, the mental choosing *M* in combination with the *relevant* antecedents of *M* (call this complex event C'), and second, the mental choosing *M* in combination with the *total* antecedents of *M* (call this complex event *C*).

To start, note that there is no in-principle problem with associating a set of token powers with either C' or *C*. After all, as we have seen, the events, entities, processes, or other goings-on at issue in debates over the status of higher-level entities are often complex and typically spatially and temporally extended; and there is no in-principle problem with assigning powers to such goings-on (reflecting, e.g., the operative laws of nature). As such, we are in position to see that the compatibilist's proper subset strategy entails that the set of relevant causal antecedents C' has a proper subset of the powers of the entire set of causal antecedents *C*. For this strategy plausibly entails that, on the one hand, every token power of C' is identical

to a token power of C, but that C' has fewer powers than C; and this is because, while C has the power to result in a choice to φ, C' does not have this power—for C' could occur in circumstances in which M (or an event of M's type) would have resulted either in a choice not to φ or in the absence of a choice to φ.

On this reconstruction of the compatibilist strategy, it would not be the seemingly free choosing itself that would be appropriately deemed Weakly emergent, but rather the seemingly free choosing in combination with the relevant proper subset of antecedents. Such a result makes sense, on a compatibilist picture. Unlike the case of a Weakly emergent qualitative feature, for example, the status of the event of choosing as genuinely free requires that the choosing be associated with causal antecedents that in the relevant sense leave the outcome up to the agent; hence on a compatibilist view, the freedom of a given act of choosing can be seen as constituted by the occurrence of a complex temporally and causally extended event, consisting in the choosing in combination with the relevant causal antecedents.

The previous considerations indicate that free will on a compatibilist account can be seen as satisfying each of the conditions in *Weak Emergence*, and so as Weakly emergent.

This result provides a new basis for addressing a frequently voiced concern about the compatibilist's strategy—namely, that identification of a given subset of causal antecedents won't make sense of how a mental choosing M could be free, since the mere presence of a subset of antecedents doesn't establish that M 'selects' or 'determines' the outcomes of the choosing.

To start, the compatibilist, like the nonreductive physicalist, will grant that M's distinctive efficacy vis-à-vis the effect at issue doesn't proceed by way of M's having a distinctive power: just as *Physical Causal Closure* blocks taking a qualitative or intentional mental event M to have a new power (that would be Strong emergence, not physicalism), so too does *Causal Determinism* block taking a mental event of choosing M to have a new power (that would be libertarianism, not compatibilism). Even so, just as the nonreductive physicalist has alternative ways of motivating the distinctive efficacy of qualitative and intentional mental events—either as tracking difference-making considerations (if the physical realizer had been slightly different, I would still have been thirsty), or as tracking a comparatively abstract psychological system of laws or causal joint—so too may the compatibilist maintain that even in the absence of new powers to 'select' or 'determine' outcomes, M may be distinctively efficacious vis-à-vis those outcomes, either in tracking difference-making considerations (if the causal antecedents of my choice had been slightly different, I would still have chosen as I did) or as tracking a distinctive broadly psychological level of causal grain. The distinctive form of causal relevance/distinctive efficacy identified by nonreductive physicalists—namely, that encoded in the schema for Weak emergence—appears to be, mutatis

mutandis, just what the compatibilist needs. Of course, compatibilism faces other challenges (a point to which I will later return), but in any case it is worth noting that the parallel to nonreductive physicalism is useful in clarifying just what the compatibilist strategy for achieving the autonomy of free will is supposed to be.

8.3 Libertarianism and Strong emergence

I next turn to libertarianism (a.k.a. 'incompatibilism'),[10] according to which both the antecedent (*Causal Determinism*) and the consequent (denying the existence of free will) of the free will conditional are false. I will argue that there is a parallel between libertarianism, on the one hand, and Strong emergentism as standardly directed at qualitative and non-intentional mental states, on the other.

8.3.1 The libertarian's new power strategy

An initial point of similarity between libertarianism and Strong emergentism is that each view rejects the broadly empirical thesis that causes trouble for the supposition that the higher-level goings-on are causally relevant in the intended sense. As Clarke and Capes (2017) put it:

> To have free will is to have what it takes to act freely. When an agent acts freely—when she exercises free will—it is up to her whether she does one thing or another on that occasion. A plurality of alternatives is open to her, and she determines which she pursues. When she does, she is an ultimate source or origin of her action. So runs a familiar conception of free will.
>
> Incompatibilists hold that we act freely in this sense only if determinism is false. (1)

Similarly, Strong emergentists maintain that an appropriate understanding of the efficacy of certain goings-on requires the falsity of *Physical Causal Closure*.

The deeper point of similarity, however, lies in the positive accounts given of the existence and causal relevance/distinctive efficacy of the higher-level goings-on at issue. Recall that the conditions in *Strong Emergence*, which takes its original inspiration from the Strong emergentist strategy for responding to the problem

[10] Here again we can distinguish between incompatibilist views which simply deny the compatibility of free will and *Determinism*, and those which moreover aim to give some positive account of the nature of incompatibilist free will; at issue in this discussion are what we might call 'positive incompatibilist' views.

of higher-level causation, require that a Strongly emergent feature have a fundamentally novel power not had (or only indirectly had, in virtue of being a sufficient precondition of the Strongly emergent feature) by the lower-level physical feature upon which it cotemporally materially depends, which novel power in turn provides a principled metaphysical basis for both the ontological and the causal autonomy of the Strongly emergent feature. As I will shortly argue, a representative range of libertarian accounts are reasonably seen as committed to free will's satisfying these conditions, and in particular to taking free will to be associated with a fundamentally novel power—namely, the power to freely choose to φ, where (as previously) a choice to φ (an outcome of an event of choosing) may be either an intention to φ, or a φ-ing.

In making this case, I'll help myself to a commonly acknowledged tripartite taxonomy of libertarian accounts, as falling under noncausal, event-causal, and agent-causal varieties. As Clarke and Capes (2017) note:

> The incompatibilist theories that have been offered fall into three main groups, depending on which type of indeterminism (uncaused events, nondeterministically caused events, agent- [or substance-] caused events) they require. Further variations among accounts concern where in the processes leading to decisions or other actions they require indeterminism and what other conditions besides indeterminism they require. (1, brackets in text)

(See also O'Connor 2005.) To fix ideas, I will primarily (though not exclusively) focus on representative instances of the accounts, as proposed by Ginet, Kane, and O'Connor, respectively. Each of these accounts can be seen as offering a different positive account of what makes a given act of choosing free in the relevant (strong, incompatibilist) sense, against the common backdrop assumption of the rejection of *Determinism*. The case for libertarian free will's being understood as Strongly emergent is most straightforward for the two causalist accounts, so I'll start with those; I'll then make a case that even so-called 'noncausalist' accounts are plausibly committed to free will's involving a (fundamentally) novel power, as Strong emergence requires.

Before getting started, I register three observations that will sometimes enter into what follows.

First, though it is more or less common ground that any variety of free will worthy of the name (whether compatibilist or incompatibilist) has to make sense of a choice to φ being in some sense 'up to the agent', the issue of such 'agential control' is often interwoven with what sort of free will might make room for moral responsibility, character, reasons, values, self-identification, conscious and unconscious intentions, and so on—'moral notions', for short. This is even more

the case in presentations or defenses of libertarian accounts, perhaps reflecting a driving motivation of such accounts to do better by way of accommodating moral notions than compatibilist or hard determinist accounts. My own view is that this mixing of the metaphysics of free will with the question of how such agency bears on moral notions is ill-advised. A more systematic approach, it seems to me, is to start by getting clear about what options the libertarian has for accommodating simple cases of seemingly free choosing, such as a case in which one considers whether to throw a piece of chalk in the air and then seemingly freely determines to do so (or not), leaving for later treatment the question of how the available options comport with the deeply complex further issues of moral responsibility, self-definition, values, and so on.

Second, and again by way of partial explanation, perhaps, of the common focus on moral notions in discussions of libertarianism, an oft-stated concern with libertarian accounts is that, insofar as they reject *Causal Determinism*, they must be committed to thinking that the outcomes of free choices are somehow indeterministically caused, via quantum or other 'chance-y' processes, in a way that would undermine agential control and hence the associated basis for accommodating moral notions. To be sure, some libertarians (e.g., Kane; see the following subsection) embrace a kind of analogy between the indeterminacy of free choice and quantum indeterminacy. It is important to realize, however, that the libertarian is under no obligation to suppose that the only alternative account to one on which choices are nomologically deterministic is one according to which such choices are nomologically indeterministic. On the contrary, the libertarian may endorse the view that agents are *nomologically transcendently* free—that is, free, in at least some of their choices, from *either* deterministic or indeterministic laws.

Third, some discussions of libertarianism take for granted that 'event' causation and 'agent' causation are distinct varieties of causation; relatedly, discussions of agent causation frequently import the assumption that such causation would be causation by a 'substance', which in turn (some complain) is problematic or in any case unusual. My own view is that there is no need to introduce a new variety of causation, much less a new substance, in order to accommodate the sort of nomologically transcendent freedom that agents possess (or should be thought to possess), on a libertarian view.

Insofar as I believe that certain methodological presuppositions of the accounts to follow are problematic, my discussion will at times depart from the letter of certain libertarian views, and advance theses that to my mind do better at metaphysically accommodating the spirit of libertarianism.

Event-causal accounts

Clarke and Capes (2017) describe event-causal libertarian accounts as follows:

> Compatibilist accounts of free will are typically event-causal views, invoking event-causal accounts of action. The simplest event-causal incompatibilist theory takes the requirements of a good compatibilist account and adds that certain agent-involving events that cause the action must nondeterministically cause it. When these conditions are satisfied, it is held, the agent exercises in performing her action a certain variety of active control (which is said to consist in the action's being caused, in an appropriate way, by those agent-involving events), the action is performed for a reason, and there remains, until she acts, a chance of the agent's not performing that action.

Standard varieties of such accounts (a.k.a. 'centered accounts') locate the indeterministic causation at issue in the immediate causal antecedents of the choice (as opposed to some prior indeterministic process leading to certain beliefs becoming salient or certain preferences being formed, which beliefs or preferences enter into the process of deliberation).

Does the indeterministic causation here involve a fundamentally novel power? Arguably, event-causal libertarians should think so, if they aim to decisively answer the so-called 'objection from luck', to which such accounts are commonly seen as vulnerable. As Clarke and Capes (2017) put the concern:

> If a decision is nondeterministically caused, and if there remains until it occurs a chance that the agent will instead (at that moment) make a different decision, then there is a possible world that is exactly the same as the actual world up until the time of the decision, but in which the agent makes the alternative decision then. There is, then, nothing about the agent prior to the decision—indeed, there is nothing about the world prior to that time—that accounts for the difference between her making one decision and her making the other. This difference, then, is just a matter of luck. And if the difference between the agent's making one decision and her instead making another is just a matter of luck, she cannot be responsible for the decision that she makes.

Consider, for example, Kane's centered causal account.[11] Bracketing certain nuances, Kane maintains that a choice to φ is one that involves a "self-forming willing"—an indeterministically caused choice or other action for which the agent is "ultimately responsible" (1996: 35).[12] Of course, not all free choices need be "self-forming"—here we see the unhelpful mixture of the more basic metaphysical question of in what libertarian free will consists with the much more complex question of how our free choices enter into self-constitution and

[11] See also Nozick 1995, Ekstrom 2001, and Franklin 2018.

[12] The primary nuance consists in allowing that a choice might be free even if causally determined, so long as it at least partly resulted from a self-forming willing.

moral responsibility. In any case, as Clarke and Capes (2017) point out, it remains unclear how appeal to indeterministically efficacious “willings” is supposed to answer the objection from luck—at least if the indeterminacy of will or effort is, as Kane unwisely suggests, analogous to the sort of indeterministic processes associated with quantum phenomena. Such an appeal not only raises the specter of the objection from luck, but also problematically suggests that free will is, like quantum goings-on, still caught in the net of nomological causation. That’s not what a so-called ‘libertarian’ should say, or so it seems to me.

At any rate, there doesn’t seem to be any reason why an event-causal libertarian such as Kane couldn’t rather maintain that the indeterminacy of agent-involving events of free choosing reflects not an analogy with (much less a basis in) indeterministic quantum goings-on, but rather that agential control involves a fundamentally novel power to choose in a way that is not a matter of either deterministic or indeterministic nomological processes. As the simple cases suggest, the novel power here is not, at least in the first instance, essentially tied to morally loaded notions involving character, morality, or the like. Rather, the power is simply the power to choose, to make a choice to intend or to act, in a way that transcends any nomological goings-on, and for whatever reason the agent happens to find compelling in the moment,[13] or indeed, for no reason at all (as in the case of an illustrative throwing of a piece of chalk). To be sure, as embodied, persons are subject to various physical and psychological limitations: to be capable of choosing to intend or to act, in a way that is neither deterministically nor indeterministically nomologically determined, does not mean that all bets, or all laws of nature, are off. Still, as persons we are capable of choosing to φ in a way that is plausibly fundamentally novel, insofar as any non-agents with which we are familiar do not have such a power to freely choose. Or so an event-causal libertarian can and should say.

I interpret Merricks (2003) as endorsing such a view. Merricks argues, as regards conscious persons, that “we should say that some of what those objects cause, in virtue of having those properties, lack microphysical causes” (110). Merricks’s claim here is compatible with an event-causal account, and he is clear that he sees exercises of free will as involving a power that lower-level physical goings-on do not have—that is, as involving a fundamentally novel power, of the sort that satisfaction of the schema for Strong emergence requires:

> Sometimes my *deciding* to do such and such is what causes the atoms of my arm to move as they do. Presumably my so deciding won’t ever be the *only* cause of their moving. There will also be a cause in terms of microphysics or microbiology, in

[13] Here an underappreciated resource might advert to the phenomenon of attention.

> terms of nerve impulses and the like. But at some point in tracing back the causal origin of my arm's moving (if it is intended), we will reach a cause that is *not* microphysical, that just is the agent's *deciding* to do something. (110)

Merricks distinguishes his view from that of the British Emergentists, on grounds that they "seem to explain *being emergent* in *epistemic* terms" (111, note 13). However, as previously discussed, that the British Emergentists sometimes characterized emergence in epistemic terms reflected their (incorrect) assumption that certain epistemic failures would track fundamental novelty; properly understood, Merricks and the British Emergentists are on the same metaphysical page. More generally, in Merricks's view, we have a form of event-causal libertarianism that plausibly satisfies the conditions in *Strong Emergence*.

Agent-causal accounts

Clarke and Capes (2017) describe agent-causal libertarian accounts as follows:

> On what are called agent-causal views, causation by an agent is held not to consist in causation by events (such as the agent's recognizing certain reasons). An agent, it is said, is a persisting substance; causation by an agent is causation by such a substance. Since a substance is not the kind of thing that can itself be an effect (though various events involving it can be), on these accounts an agent is in a strict and literal sense an originator of her free decisions, an uncaused cause of them. This combination of indeterminism and origination is thought to capture best the idea that, when we act freely, a plurality of alternatives is open to us and we determine, ourselves, which of these we pursue, and to secure the kind of freedom needed for moral responsibility. (23)

While in some cases talk of an object or other particular entity's causing an effect can be seen as shorthand for talk of the object's having some feature that produced the effect (as when, e.g., 'the rock broke the window' is more specifically understood as 'the rock's having momentum M at the point of contact with the window caused the window's shattering'), the agent-causal libertarian denies that, for the outcomes of free choices, such a reduction (e.g., to the agent's recognizing certain reasons) is available.

Concerns about agent causation often stem from concern about causes being substances rather than events. One might reasonably deny, however, that making sense of agent-causation requires thinking of agents as 'substances' or positing an associated new form of causation, as opposed to thinking of agents as objects of a distinctive variety, no more problematic than the distinctive kinds of objects (cells, planets, plants, and so on) posited by the various special sciences, which are also supposed to be capable of causing various effects. No doubt such special-science

objects cause these effects in virtue of certain of their properties, as per usual. But that needn't be a problem for the agent-libertarian: if the properties relevant to the exercise of free choice are along lines of 'having nomologically transcendent free will', then there is no threat to the supposition that agents are capable of causing the outcomes of their acts of free choosing. Again, if there is a salient causal difference between agents and other kinds of objects, it has to do with whether or not a given instance of causation is law-governed (deterministically or indeterministically) or not; and this difference can be registered without taking agents to be substances or to exercise a new form of causation.

In any case, on agent-causal accounts there is a clear sense in which free will involves the exercise of a fundamentally novel power. O'Connor (2009*a*) is explicit on this score, saying that

> [A]n adequate account of freedom requires, in my judgment, a notion of a distinctive variety of causal power, one which tradition dubs "agent-causal power". (191)

> The familiar considerations [against locating free will in indeterminacy] lead certain philosophers to conclude that the kind of control necessary for freedom of action involves an ontologically primitive capacity of the agent directly to determine which of several alternative courses of action is realized. (193)

O'Connor is also explicit that the posit of such a power can be understood in terms of emergence rather than substance dualism:

> While the tidiness of substance dualism has its appeal, it is in fact optional for the metaphysician who believes that human beings have ontologically fundamental powers (whether of freedom or consciousness or intentionality). For we may suppose that such powers are […] ontologically emergent powers, ones that are at once causally dependent on microphysically-based structural states and yet ontologically primitive, and so apt to confer ontologically primitive causal power. (191)

What is the power at issue? As desired, the fundamentally novel power is one which allows one to choose to act, or to intend to act, in one way rather than another:

> One important feature of agent-causal power is that it is not directed to any particular effects. Instead, it confers upon an agent a power to cause a certain type of event within the agent: the coming to be of a state of intention to carry out some act, thereby resolving a state of uncertainty about which action to undertake. (195)

> Agent causation is precisely the power to directly determine which of several causal possibilities is realized on a given occasion. [...] [T]he view posits a fundamental, irreducible power of agents to form intentions. (211–12)

The posit of such a novel fundamental power is supposed to provide a basis for responding to the luck objection, in virtue of positing "a kind of single-case form of control by means of which the agent can determine what happens in each case" (213). For present purposes, what is crucial is that O'Connor's version of an agent-causal form of libertarianism clearly satisfies the conditions in *Strong Emergence*.

Noncausal accounts

Clarke and Capes (2017) describe noncausal libertarian accounts as follows:

> Some incompatibilist accounts require neither that a free action be caused by anything nor that it have any internal causal structure. Some views of this type require that a free action be uncaused; others allow that it may be caused as long as it is not deterministically caused. Since any such account imposes no positive causal requirement on free action, we may call views of this type "noncausal".

As O'Connor (2005) puts it, on a noncausal account, free will is taken to be "entirely noncausal in character and is instead a consequence of intrinsic, noncausal features of the choice itself". There are several different conceptions of the intrinsic feature at issue. On Ginet's version (see, e.g., Ginet 1990 and 2002), this feature is an "actish phenomenal quality," which he describes (1990, 13) as its seeming to the agent as if they are directly producing, making happen, or determining the event that has this quality. On McCann's (1998) version, the intrinsic feature at issue is 'intrinsic intentionality'. As Clarke (2003) describes McCann's view:

> [I]n making a decision, McCann maintains (1998, 163), one intends to decide—indeed, one intends to decide exactly as one does (e.g., when one decides to A, one intends to decide to A). One's so intending, though intrinsic to the decision, is not a matter of the content of the intention that is formed in deciding; nor is it a matter of one's having any further intention in addition to that formed in making the decision. Rather, McCann holds, it is a matter of a decision's being, by its very nature, an act that the agent means to be performing. (18)

Finally, on Stump's (1999) version, the intrinsic feature is one along Thomistic lines, as involving an act of will or intellect:

> What is essential to moral responsibility on Aquinas's view is that a person be the ultimate source of what she does, that her intellect and will be the ultimate causes of her acts. By 'ultimate cause' here, I mean that there is nothing which is prior to that person's acts of intellect and will and which causally determines her intellect and will to be in the states in which they are. If we can trace a causal chain of any sort backward from an agent's act, then the causal chain must originate only in acts of her will and intellect. That is, for any act which the agent does, if there is any causal chain at all of which the act is the effect, then the causal chain must have a first or ultimate cause, and that ultimate cause cannot be anything other than an act of the agent's own will or intellect. (414)

I will now argue that, on plausible construals of each of these variations on the noncausal theme, the sense in which free will is 'noncausal' pertains only to the purported failure of free choices to themselves be effects of prior (deterministic or indeterministic) causes. This much is compatible, however, with free choosings on such an account being 'causal' in the sense of themselves having effects, and associated powers. To prefigure: this result, coupled with the supposition that the powers themselves are fundamentally novel relative to the powers of dependence base goings-on, supports taking free will on a noncausal account to conform to the conditions in Strong Emergence.

Consider, to start, Ginet's (1990) account, according to which free will involves "an actish phenomenal quality". As Clarke (2003) observes, one might complain that such a conception is compatible with an agent's not really having any "active control":

> Whatever the correct characterization of this phenomenal quality, the mere feel of a mental event—the way it seems to the individual undergoing it—although it may be a (more or less reliable) sign of active control, cannot itself constitute the agent's exercise of such control (cf. O'Connor 2000, 25–26). To hold that it does is to render the exercise of active control wholly subjective (nothing more than the way things seem), and this is to greatly diminish the significance of active control. (20)

Ginet is within his rights, however, to insist that just because the quality at issue is phenomenal, it does not follow that this renders the associated act of choosing "wholly subjective" or "nothing more than the way things seem". For notwithstanding that actish phenomenal qualities are noncausal in not being caused, this is compatible with such qualities having powers to produce effects—most saliently, the outcomes of acts of choosing.

Next, consider McCann's (1998) view, according to which noncausal free will involves intrinsic intentionality, and where, more specifically, an act of libertarian

choosing is one which is characterized as such by the presence of an intention to decide, as "a matter of a decision's being, by its very nature, an act that the agent means to be performing" (Clarke 2003, 18). Here again, and granting that both the act of choosing and the outcome of the choice are noncausal in not being appropriately seen as effects of prior causes, it remains that an act of choosing is (also) causal precisely in that it involves the exercise of a power to choose, or decide.

Finally, consider Stump's (1999) Thomistic view, according to which libertarian free will is noncausal in the sense that, when a person chooses, "there is nothing which is prior to that person's acts of intellect and will and which causally determines her intellect and will to be in the states in which they are" (414). Here a given act of choosing may, after all, be caused—so long as those causes are other acts of intellect or will. Most importantly for present purposes, however, this Thomistic view is compatible with taking such acts of intellect or will to be causal in the sense of themselves having powers to cause or contribute to causing certain effects—most saliently, in a case of free choosing, the outcome of the act of choosing.

On a representative range of noncausal accounts of libertarian free will, then, the characterization of acts of choosing as 'noncausal' reflects that such acts are not the effects of (non-will-involving) causes prior to the acts. It remains that such acts are causal at least in having powers to cause, at a minimum, an outcome of the choosing, and moreover to contribute to causing any effects associated with the outcome of the choice (e.g., the reaching for a glass of water). This distinctive causal asymmetry—free choices are uncaused, but capable of causing—is reasonably interpreted as suggesting that on noncausal libertarian accounts, acts of free choosing involve fundamentally novel powers, as Strong emergence requires.

8.4 Is free will either Weakly or Strongly emergent?

I turn now to considering whether there is actually any free will of either Weak or Strong emergent varieties.

8.4.1 Is there compatibilist (Weakly emergent) free will?

As discussed in §8.2, the compatibilist is plausibly seen as implementing a proper subset strategy, on which the freedom of an act of choosing is ultimately a matter of the act's being associated with a proper subset of the complete and actual causal antecedents of the choosing; and as I argued above, this strategy can in turn be plausibly interpreted as involving certain complex events' more specifically

satisfying the conditions in *Weak Emergence*. Given all this, what are the prospects for there actually being free will of the Weakly emergent variety?

The prospects are good. Though free choices are not taken to be part of a higher-level system of laws on either compatibilist or libertarian accounts, a compatibilist account is one manifesting the usual Weak emergentist characterization of special-science goings-on as comparatively insensitive to lower-level physical details, in the sense that an agent's reasons for action in a given case float free of many such details (and in particular, are sensitive only to facts about 'relevant' causal antecedents). Since our deliberations and associated acts of choice clearly are insensitive to many microphysical details, then given that free will is understood along compatibilist (Weak emergentist) lines, there is good reason to think that such free will actually exists, and moreover is abundant. Indeed, even libertarians grant that there is actually free will—or as they would prefer to call it, 'active control'—of a compatibilist variety. Hence Clarke (2003) says:

> We make decisions and act even if determinism is true; we are thus unlike puppets. And unlike agents who are not persons, we can still act on the basis of our appreciation of practical reasons, including moral reasons. We are also unlike prisoners, in that we can generally go where we want to go. Further, most of us most of the time act quite free from coercive threats and compulsive desires. We are never subject to the direct manipulation of our brains by malevolent neuroscientists. All of this is good, and these goods do not require indeterminism. There is, then, a valuable variety of active control that we can have and exercise even if determinism is true. However, this compatibilist variety of active control falls short of free will. (8)

8.4.2 Is there libertarian (Strongly emergent) free will?

Notwithstanding that there is presumably plenty of what the compatibilist counts as free will, I am inclined to agree with those, like Clarke, who think that compatibilist accounts are ultimately unsuccessful in accommodating the core phenomenal and intentional aspects of free will, according to which a freely choosing agent feels, from the inside, to be causally determinative of the outcome in a way transcending any nomological goings-on, whether deterministic or indeterministic. Even those who aim to reject the appearances as genuine admit as much. Hence Caruso (2012), a determinist, says:

> A major part of the folk psychology of free will is the belief that *our conscious intentions cause action*. As Patrick Haggard and Benjamin Libet write, "Most of us navigate through our daily lives with the belief that we have conscious free will: that is, we have conscious intentions to perform specific acts, and those intentions

> can drive our bodily actions, thus producing a desired change in the external world" (2001, 47). This commonsense intuition plays a major role in our sense of free will and is essential to the *up-to-me-ness* that we associate with free will. It is also well supported by phenomenology. In normal cases of voluntary behavior, we experience a conscious intention before the onset of action and naturally take the former to be the cause of the latter. When I switch on my reading lamp, for example, I feel as though it is *I*, my conscious self, that controls the movements of my arms and hands through the conscious formation of goals, intentions, and decisions. (189)

Two points about this sort of pretheoretical 'take' on our capacity for free will are worth noting. First, note that the phenomenal and intentional motivations for thinking that our choices are at least sometimes nomologically transcendently free don't turn on the choice at issue being directed at anything morally or otherwise substantive.[14] I can experience myself as seemingly transcendently free even if all that is at issue is whether or not to throw a piece of chalk into the air.[15] As such, and even granting that inquiring into exactly how free will (assuming it exists) intersects with notions such as reasons, character, and moral responsibility, for present purposes we can restrict our focus to what seems to me to be the prior question of whether we are ever in position to act in ways that transcend the nomological net. Second, note that this much is already enough to block the too-quick strategy of appealing to indeterministic quantum or other laws as a basis for such freedom. This strategy is often rejected on grounds that an appeal to indeterministic laws would render the outcomes of choosings subject to random processes rather than reasons; but for present purposes it is enough to reject this strategy to note that indeterministic laws are still laws, solidly within the net of nomological goings-on, in which case appeal to such laws cannot provide a realistic metaphysical basis for the seeming experience of nomologically transcendent free will.

Now, if the appearance of nomologically transcendent free will is to be taken as genuine, free will so understood is not properly accommodated via the usual Weak emergentist approach to higher-level special scientific goings-on, for even if it is correct (as the compatibilist maintains) that our acts of seemingly free choosing are insensitive to certain (antecedent) micro-level details, it is not this insensitivity that constitutes nomologically transcendent free choice, or the experience of such. Indeed, once it is appreciated that compatibilist accounts of free will are

[14] This observation is registered in views on which freedom is not itself sufficient for moral responsibility; see Clarke 1992.

[15] Hence I would disagree with Clarke's (2003)'s claim that "For an agent to act with free will, she must be able to regard some considerations as reasons for action" (16), at least if the 'reason' has to contain more content than the choice in itself, as in 'I hereby choose to throw the piece of chalk into the air'.

implementing a variation on the usual Weak emergentist theme, such that any token power associated with (a complex event having as a part) an act of choosing ends up being identical to a token power associated with a lower-level physical event—to be sure, a highly complex, temporally extended physical event, but no matter—the usual concerns with a compatibilist approach are thrown into high relief. It may be that agents on a compatibilist approach are distinctively efficacious, in ways that complex systems, ordinary objects, and other special-science goings-on are distinctively efficacious, in having distinctive power profiles that track difference-making considerations (if I had been differently molecularly configured, I would still have chosen to go to Pasternak for brunch) or comparatively abstract causal joints. But this form of distinctive efficacy is not of the transcendent variety that is core and crucial to our phenomenal and intentional experience of agency.

What seems to be required for realistic accommodation of this experience is that associated acts of free choosing involve a fundamentally novel power which, as libertarians standardly maintain, transcends any nomological net. And as previously discussed, a fundamentally novel power of event-causal, agent-causal, or 'noncausal' variety will do for purposes of such accommodation.

Is there any libertarian (Strongly emergent) free will, so transcendently understood? To start, as with other purported Strongly emergent phenomena, there is a case to be made that the phenomenon of libertarian free will admits of empirical confirmation (or disconfirmation), at least in principle, by attention to whether the phenomenon involves any apparent violations of conservation laws. After all, even if libertarian free will involves a fundamentally novel power, and even if such a novel power transcends not just the nomological net of lower-level physical goings-on, but indeed any nomological net, so long as such a power is at least partly determinative of a given effect, it will presumably involve some transfer of energy or other conserved quantity; and if such a transfer involves some fundamentally novel interaction, then this would provide an in-principle means of empirically verifying the hypothesis that free will is Strongly emergent.

Independent of this route to empirical verification, however, the fact that we have direct introspective access to the phenomenon of seemingly nomologically transcendent free will provides the basis for a new argument for there actually being libertarian—Strongly emergent—free will, as follows:

1. We experience ourselves as seeming to freely choose, in ways transcending any nomological (deterministic or indeterministic) goings-on.
2. In the absence of good reasons to think that our experience of nomologically transcendent free will cannot be taken at face value, we are entitled to take this experience at realistic face value.
3. There are no good reasons to think that our experience of transcendent free will cannot be taken at realistic face value.

∴ We are entitled to take our experience of nomologically transcendent free will at realistic face value.

The argument is valid, and premise 1 is clearly true: again, even non-libertarians agree that we *seem* to freely choose, in ways transcending any nomological net, on at least some occasions. Premise (2) also seems reasonable: if we have clear experience of some seeming phenomenon, we need good reason not to take that experience at realistic face value. In what follows, then, I'll focus on defending premise (3) against certain salient empirical results purporting to show that the phenomenon of free will is in some sense an illusion.

The empirical case against libertarianism

Recent results in empirical neuroscience have frequently been taken to show that seeming experiences of free will as nomologically transcendent cannot be taken at face value. Hence Caruso (2012) claims:

> Although phenomenology supports this commonsense belief [in nomologically transcendent free will], empirical evidence in neuroscience now seriously questions it. In fact, a growing number of theorists now conclude that *conscious will*—in the sense of consciously initiated action—is incompatible with the evidence of neuroscience [...] Much of the contemporary case for this conclusion is derived from the experimental work of Benjamin Libet and his colleagues. [...] My thesis will be that the empirical results from neuroscience do in fact reveal that conscious will is an illusion—at least in the cases we can currently study empirically. (189)

In what follows I focus on the 'Libet cases' which pose the most serious challenge to taking our seeming experience of transcendent free will at face value.[16] These studies aim to compare the self-reported time of occurrence of a certain conscious choice to produce a certain physical behavior with the time of occurrence of certain unconscious brain states associated with the production of the behavior. The concern raised for nomologically transcendent free will is that the evidence gathered in these studies has been interpreted as suggesting that the unconscious initiation of the physical behavior occurs *prior* to the time of the supposed choosing, such that the supposition that our free choices are determinative of our

[16] See O'Connor (2009*b*) for discussion of a number of other empirical results that some have seen as problematizing taking the seeming nomological transcendence of free choice as genuine; these include, e.g., cases where subjects of induced behaviors confabulate instances of their agency in a post-hoc way, cases where the outcomes of supposedly free choices (about which index finger to move) are influenced by external stimulation of the subject's brain, cases of subjects who report a feeling of agency concerning distal outcomes subsequent to being instructed to have certain negative thoughts, and cases where subjects engage in seemingly voluntary action unaccompanied by any feeling of agency (as in 'alien hand syndrome'). O'Connor compellingly argues that none of these results poses a serious challenge to the libertarian supposition; I direct the interested reader to his discussion.

physical actions is illusory. O'Connor (2009*b*) describes the original setup and interpretation of results as presented in Libet 1999:

> Libet devised a study in which people are asked to wiggle their finger within a short interval of time (thirty seconds or so). The experimenter instructs them to do so whenever they wish—though spontaneously, not by deciding the moment in advance. Throughout, they are to watch a special clock with a very fast-moving dial (a beam of light) and note its location at the precise moment at which they felt the "urge" or "wish" to move the finger. During the experiment, a device measures electrical activity on the agent's scalp. Libet discovered that a steady increase in this activity (dubbed the "readiness potential," or RP) consistently preceded the time the agents cited as when they experienced the will to move. By averaging results over hundreds of experiments, Libet determined that the RP preceded the "experience of will" by an average of some 400 milliseconds, a significant interval in the context of neural activity. Libet and others concluded from this result that "conscious will" is not the initiator of voluntary acting but instead a consequence of an unconscious physical process that also (and according to some hypotheses, independently) triggers the action. (176)

Do experimental studies of this sort really establish that nomologically transcendent free will is an illusion? This question continues to be hotly debated, but so far as I can tell, the answer is clearly 'no'. Others—notably, O'Connor (2005b) and Mele (2009) have offered very detailed discussions of the relevant studies, noting in particular that these admit of plausible alternative interpretations which do not have anti-libertarian import. Here I'll briefly canvass two of these interpretive options, and offer a new one of my own, which reflects one of the characteristic features of metaphysically emergent phenomena.

To start, as O'Connor notes, the setup is one where the agent has already decided to perform a specific action, and while the focus of the experiment is supposedly on when the agent's choice to perform that action occurs, both the antecedent decision and the restriction on timing render the setup sufficiently nonstandard that one might reasonably deny taking the results to suggest anything general about the status of nomologically transcendent agency. More to the interpretive point, O'Connor observes that the instructions to the subject to introspect and wait for an unplanned "urge" to occur would likely encourage a passive posture in which "having decided that one will move, one looks for the *urge* to do so in order to act upon it", which in turn motivates an alternative interpretation of the results according to which "such a preformed intention to act upon the right internal 'cue' initiates an unconscious process that promotes the occurrence (or perhaps *evolution*) of a conscious state of desire or intention that is not actively formed" (182).

Such an interpretation is consonant with another study by Libet, which confirms that in such an experimental context, the subject has 'veto' power over the antecedent 'urge' in question. As Libet (2002) himself observes:

> [T]he conscious function still had enough time to affect the outcome of the process; that is, it could allow the volitional initiative to go to completion, it could provide a necessary trigger for the completion, or it could block or veto the process and prevent the act's appearance. There is no doubt that a veto function can occur. (292)

One might thus interpret the antecedent brain activity in the original setup as simply setting in train a deliberative process having as its 'default' object the performance of a certain physical action, where the act of free choosing consists *either* in a decision to allow the process to continue to completion *or* in a decision to block the process. A second, more explicitly causal, interpretation is found in Mele's (2009) suggestion that the ability of subjects to exercise veto power renders it "much more likely that [the prior brain activity] is a potential cause of a proximal intention or decision rather than a proximal intention or decision itself" (51). Each of these alternative interpretations is compatible with nomologically transcendent free will.

Finally, I suggest that a third interpretive option should be put on the table—namely, one according to which the intention to choose and the brain activity are cotemporally initiated, but where it takes just a bit of time for this fact to consciously register as a complete thought in the agent's mind. Thinking, and presumably also choosing, takes time. As such, a very small lag—less than half a second—between initiation and conscious recognition would be a natural concomitant of our mental decision-making processes, again compatible with nomologically transcendent free will. More generally, Libet's assumption that "In the traditional view of conscious will and free will, one would expect conscious will to appear before, or at the onset, of the RP [Readiness Potential], and thus command the brain to perform the intended act" (1999, 49) appears to reflect an unmotivated understanding of transcendent free will as instantaneous and indeed as in some sense floating free of underlying physical processes.

Here, it seems to me, attention to the characteristic features of metaphysical emergence can prove useful. For first, the libertarian qua Strong emergentist will assume that mental features are cotemporally materially dependent on physical features; but second, as I have previously emphasized, the cotemporal dependence base may be temporally extended. As such, one might reasonably maintain, as an available interpretation of Libet's results, that the occurrence of a given Readiness Potential would be not prior to the event of choosing, either as a kind of 'default' unconscious deliberative process (as O'Connor suggests) or as a prior cause of the event of choosing (as Mele suggests), but rather as a *part* of the temporally extended dependence base for an event of free choosing.

There are thus at least three ways to resist taking Libet cases to undercut our seeming experience of nomologically transcendent free will. I conclude that in the absence of any more compelling route to undercutting this experience—which experience, again, is plausibly understood as involving satisfaction of the *New Power Condition*—and given the empirically supported supposition that acts of

free choosing (like other mental events) are cotemporally materially dependent on lower-level goings-on, we are entitled to take our experience at realistic libertarian face value as an actual case of Strong emergence.

8.5 Concluding remarks

Let's sum up. As Bernstein and I argue in our 2016, there is an important and theoretically powerful connection between the problem of mental causation and the problem of free will—namely, each can be seen as specific cases of a general 'problem of mental quausation'—the problem of how a mental event (or other higher-level feature) can be efficacious *qua* the type of event it is, given the open possibility of certain theses which threaten to undermine, one way or another, the supposed causal relevance/distinctive efficacy of the event.

This deep connection suggests certain parallels between positions in the corresponding debates. Here I have argued that compatibilists, like non-reductive physicalists, implement a strategy entailing satisfaction of the *Proper Subset of Powers Condition*, and that libertarians, like Strong emergentists generally, implement a strategy entailing satisfaction of the *New Power Condition*. Coupled with the assumption of cotemporal material dependence, it follows that compatibilist free will (if such exists) is Weakly emergent, whereas libertarian free will (if such exists) is Strongly emergent.

Finally, I have considered whether free will is actually either Weakly or Strongly emergent. Free will on a compatibilist view, I have argued, is easy to come by, as another case-in-point of the usual Weak emergentist understanding of higher-level features as comparatively abstract or insensitive to lower-level physical details. This result is a double-edged sword, however, for unlike cases of higher-level phenomena which seem amenable to Weak emergentist treatment, including complex systems, classical ordinary objects, and qualitative perceptual mental states, such insensitivity to lower-level detail seems beside the point of generally accommodating free will, which in at least some manifestations appears to have more to do with an agent's ability to nomologically transcend, as opposed to abstract from, any lower-level physical goings-on. As I have argued, however, our introspective experience of seemingly nomologically transcendent free will itself (in combination with the usual assumption of cotemporal material dependence) provides good reason to think that we have free will of a libertarian, Strongly emergent variety; and the neuroscientific reasons for rejecting this experience as illusory are uncompelling. Contrary to common assumption, it is libertarian free will, not subjective or qualitative experience, that provides the best case of a Strongly emergent phenomenon.

I conclude that there is actually free will of both Weak and Strong varieties—a nice result, given the importance of free will as a basis for personal and moral autonomy, and one which, to my mind, provides a fitting closing indication of the importance of metaphysical emergence to our understanding not just of the world, but of ourselves.

9

Metaphysical emergence: next steps

This book began with two key questions, inspired by the target cases of special-scientific and artifactual entities and features, as seemingly cotemporally materially dependent on, yet ontologically and causally autonomous with respect to, lower-level micro-configurations and features. What is metaphysical emergence, more precisely? And is there any such emergence, in principle and moreover in actual fact? We now have answers to both questions in hand.

First, what metaphysical emergence is, more precisely, is encoded in the schemas for Weak and Strong emergence, tracking varieties of emergence that are physically acceptable and physically unacceptable, respectively, on the assumption that the dependence base goings-on are ultimately physical. The conditions in the schemas—*Weak Emergence* and *Strong Emergence*, respectively—capture what is core and crucial (and when suitably filled in, necessary and sufficient) for each form of emergence, in appropriate and illuminating fashion. Moreover, given that metaphysically emergent entities and features must be causally autonomous, the two schemas exhaust the available options. For they reflect that (as attention to the problem of higher-level causation confirms) there are two and only two routes to distinctive higher-level efficacy: either the higher-level entity or feature has more powers, or it has fewer powers, than its dependence base, on a given occasion.

Second, there is metaphysical emergence, both in principle and as a matter of actual fact. Notwithstanding the many attempts to problematize the forms of emergence at issue in the schemas, these attempts ultimately fail. Each form of emergence is viable: coherent, metaphysically substantive, naturalistically acceptable, such as to avoid problematic causal overdetermination, and more generally such as to accommodate and illuminate the prima facie appearances of metaphysical emergence, as per our guiding methodology. Moreover, each form of emergence is arguably actually instantiated, with certain complex systems, ordinary objects, qualitative mental states, and compatibilist free will being Weakly emergent, and libertarian (nomologically transcendent) free will being Strongly emergent.

This is not the end of the journey, of course. To start, there are many candidates for metaphysically emergent phenomena that I have not considered in any detail here, including quantum entanglement (see Humphreys 1997, Silberstein and McGeever 1999), the fractional quantum Hall state (see Morrison 2006, Lancaster and Pexton 2015), molecular structure (see Hendry 2010 and 2017), polymers (see McLeish *et al.* 2019), biological systems (see Mossio *et al.* 2013, McLeish 2017,

Metaphysical Emergence. Jessica M. Wilson, Oxford University Press (2021). © Jessica M. Wilson.
DOI: 10.1093/oso/9780198823742.003.0009

Santos forthcoming), brain dynamics (see Thompson and Varela 2001 and Thompson 2007), collective individuals (see Bouchard and Huneman 2013), ecological systems (see Bergandi and Blandin 1998, Levin 2005, and Smith 2006), and beyond. And the status of many phenomena as Strongly emergent, including some of those we have previously considered, remains an open empirical question, contingent on as-yet-unconducted experiments establishing that (on the model of the discovery of the weak nuclear interaction) one or more fundamental interactions come into play only under certain comparatively complex circumstances. It is also worth noting that, notwithstanding that the present investigation has been directed at identifying what it would be for there to be metaphysical emergence of the sort motivated by the target cases, as combining cotemporal material dependence with ontological and causal autonomy, there may well be other cases of purported emergence—e.g., spacetime (see Huggett and Wüthrich 2013 and Lam and Wüthrich 2018 for discussion) or numbers (perhaps along lines discussed in Maddy 1997) which (depending on whether the cases really are ones of metaphysical emergence) require an alternative to or extended understanding of the schemas for metaphysical emergence presented here.[1] However these and other philosophical and empirical investigations into metaphysical emergence play out, my hope is that the present project will provide a useful base of operations going forward.

I want to close with some methodological observations, that point towards some other ways in which the present project might be profitably extended. At several junctures along the way, we saw that one or other revisionary (e.g., reductionist or anti-realist) account of the nature of seemingly higher-level entities or features reflected the supposition that some such approach is required for a naturalistically acceptable resolution of the problem of higher-level causation. A related supposition was tacitly operative in Thomasson's (2007) claim (discussed at the end of Ch. 6) that the best response to causal overdetermination-based concerns about ordinary objects is to maintain that investigations into the nature and existence of such objects should proceed in some way different from investigations into the nature and existence of special-scientific goings-on, since (she assumes) the embrace of a unified methodology would likely entail acceptance of some or other revisionary account of ordinary objects. Supposing so, she suggests, better to

[1] There is a great deal of interest to say about such cases, which must await a future occasion. For the case of spacetime, I will here just note that insofar as the purported dependence base is understood as itself not concretely located (if location presupposes spacetime) but rather in some sense abstract, an alternative understanding of the relation at issue might be as involving instantiation rather than (anything answering to) metaphysical emergence. Alternatively, if the base is somehow concretely located in such a way as to functionally implement certain features of spacetime (along lines suggested, e.g., by Knox 2013 and 2018), then it might be that references to 'cotemporal' dependence in the schemas for emergence could be extended to refer to whatever in the dependence base plays the associated role.

take for granted that ordinary objects exist and have the natures that we usually attribute to them, even at the price of turning metaphysical investigation into a branch of linguistic or conceptual anthropology.

A moral of the results argued for here, however, is that we are not forced to embrace either a revisionary metaphysics or a revisionary methodology of metaphysics—at least for all the problem of higher-level causation shows. For the concerns associated with the problem of higher-level causation can be answered, whether as attaching to special-science entities or to ordinary objects (and their features), and correspondingly there is no clear pressure from this direction for treating ordinary objects differently from (other) special-science entities.

That said, additional concerns have been raised about the posit of macro-entities and features, understood as cotemporally materially dependent on, yet autonomous with respect to, lower-level entities and features (see Paul 2010 for an overview). One of these is the 'problem of the many' (see Unger 1980), according to which the intimate dependence of, e.g., a cat—say, Tibbles—on its constituting matter, coupled with the seeming fact that there are many candidate 'cat-constituters' in the vicinity of Tibbles, poses a difficulty for the usual assumption that there is just one cat on the mat. Another is the puzzle of the statue and the clay (see Gibbard 1975), associated with the 'grounding problem' (see deRosset 2011): given the intimate relation between a statue and its constituting matter, how is it that these can be associated with different persistence conditions, such that, e.g., the clay, but not the statue, can survive squashing, and the statue, but not the clay, can survive the loss of a significant part?

These problems have also been taken by some to support revisionary metaphysical or methodological approaches to the phenomena at hand. A natural next question, then, is whether the powers-based schemas for Weak and Strong emergence, or some variation on a powers-based theme, might provide the basis for non-revisionary resolutions of problems besides that of higher-level causation. For example, perhaps the key to realistic accommodation of the differing persistence conditions of the statue and the clay lies in appreciating that, on any given occasion of constitution, the token powers of the statue and the token powers of the clay overlap—perhaps not by way of one collection being a proper subset of the other, but rather by each sharing some but not all powers with the other. Here—to venture even further along the speculative branch—there may be a reciprocal relation between the powers that are in some sense characteristic of an object and the changes through which it may persist. For example, it might be that the ability of the clay, but not the statue, to survive being squashed reflects that while the clay (characteristically, as a matter of its nature) has powers associated with the determinable feature *being shaped*, it does not have all the powers associated with *being statue-shaped* (or at least does not have those powers in the same direct way as the statue it constitutes); similarly, it may be that the ability of

the statue, but not the clay, to survive the loss of a significant part, reflects that while the statue (characteristically, as a matter of its nature) has powers associated with *being statue-shaped*, it does not have powers associated with certain specific determinates of this feature—those powers are had, at best, by certain temporal parts of the statue.

All this is speculative but to my mind tantalizing food for future thought. Time will tell, but it may well be that attention to the broadly mereological relationships between sets of powers associated with entities or features can provide the key to realistic and illuminating accommodation of other aspects of higher-level reality, beyond metaphysical emergence.

Bibliography

Aizawa, Kenneth and Carl Gillett, 2009. "The (Multiple) Realization of Psychological and Other Properties in the Sciences". *Mind and Language*, 24:181–208.

Akiba, Ken, 2004. "Vagueness in the World". *Noûs*, 38:407–29.

Alexander, Samuel, 1920. *Space, Time, and Deity*. London: Macmillan.

Alter, Torin, 1998. "A Limited Defense of the Knowledge Argument". *Philosophical Studies*, 90:35–56.

Alter, Torin, 2020. "Physicalism Without Fundamentality." *Erkenntnis*. https://doi.org/10.1007/s10670-020-00285-6.

Anderson, P. W., 1972. "More is Different". *Science*, 177:393–6.

Antony, Louise M., 2003. "Who's Afraid of Disjunctive Properties?" *Philosophical Issues*, 13:1–21.

Antony, Louise M. and Joseph M. Levine, 1997. "Reduction with Autonomy". *Philosophical Perspectives*, 11:83–105.

Armstrong, David M., 1978. *Universals and Scientific Realism, Vol II: A Theory of Universals*. Cambridge: Cambridge University Press.

Audi, Paul, 2012*a*. "Grounding: Toward a Theory of the In-Virtue-of Relation". *Journal of Philosophy*, 109:685–711.

Audi, Paul, 2012*b*. "Properties, Powers, and the Subset Account of Realization". *Philosophy and Phenomenological Research*, 84:654–74.

Auyang, Sunny, 1999. *How is Quantum Field Theory Possible?* Oxford: Oxford University Press.

Ayer, Alfred J., 1954. "Freedom and Necessity". In his *Philosophical Essays*, 271–84. London: Macmillan. Reprinted in Watson 1982, 15–23.

Baker, Lynne Rudder, 1993. "Metaphysics and Mental Causation". In John Heil and Alfred Mele, editors, *Mental Causation*, 75–96. Oxford: Clarendon Press.

Baltimore, Joseph A., 2013. "Careful, Physicalists: Mind–Body Supervenience Can Be Too Superduper". *Theoria*, 79:8–21.

Barnes, Elizabeth, 2006. *Conceptual Room for Ontic Vagueness*. Ph.D. thesis, University of St. Andrews.

Barnes, Elizabeth, 2010. "Ontic Vagueness: A Guide for the Perplexed". *Noûs*, 44:601–27.

Barnes, Elizabeth, 2012. "Emergence and Fundamentality". *Mind*, 121:873–901.

Barnes, Elizabeth and J. R. G. Williams, 2011. "A Theory of Metaphysical Indeterminacy". In Karen Bennett and Dean W. Zimmerman, editors, *Oxford Studies in Metaphysics volume 6*, 103–48. Oxford: Oxford University Press.

Batterman, Robert W., 1998. "Why Equilibrium Statistical Mechanics Works: Universality and the Renormalization Group". *Philosophy of Science*, 65:183–208.

Batterman, Robert W., 2005. "Critical Phenomena and Breaking Drops: Infinite Idealizations in Physics". *Studies in History and Philosophy of Science Part B*, 36:225–44.

Batterman, Robert W., 2000. "Multiple Realizability and Universality". *British Journal for the Philosophy of Science*, 51:115–45.

Batterman, Robert W., 2002. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford: Oxford University Press.

Baysan, Umut, 2014. *Realization and Causal Powers*. Ph.D. thesis, University of Glasgow.

Baysan, Umut, 2016. "An Argument for Power Inheritance". *Philosophical Quarterly*, 383–90.

Baysan, Umut and Jessica M. Wilson, 2017. "Must Strong Emergence Collapse?" *Philosophica*, 91:49–104.

Bealer, George, 1997. "Self-Consciousness". *Philosophical Review*, 106:69–117.

Bealer, George, 2002. "Modal Epistemology and the Rationalist Renaissance". In Tamar Gendler and John Hawthorne, editors, *Conceivability and Possibility*, 71–126. New York: Oxford University Press.

Bedau, Mark A., 1997. "Weak Emergence". *Philosophical Perspectives 11: Mind, Causation and World*, 11:375–99.

Bedau, Mark A., 2002. "Downward Causation and the Autonomy of Weak Emergence". *Principia*, 6:5–50.

Bedau, Mark A., 2008. "Is Weak Emergence Just in the Mind?" *Minds and Machines*, 18: 443–59.

Bedau, Mark A., 2010. "Weak Emergence and Context-Sensitive Reduction". In Antonella Corradini and Timothy O'Connor, editors, *Emergence in Science and Philosophy*, 6–46. New York: Routledge.

Beebe, James R., 2009. "The Abductivist Reply to Skepticism". *Philosophy and Phenomenological Research*, 79:605–36.

Bennett, Karen, 2003. "Why the Exclusion Problem Seems Intractable and How, Just Maybe, to Tract It". *Noûs*, 37:471–97.

Bennett, Karen, 2015. " 'Perfectly Understood, Unproblematic, and Certain': Lewis on Mereology". In Barry Loewer and Jonathan Schaffer, editors, *A Companion to David Lewis*, 250–61. Oxford: Wiley Blackwell.

Bennett, Karen, 2017. *Making Things Up*. Oxford: Oxford University Press.

Bergandi, Donato and Patrick Blandin, 1998. "Holism vs. Reductionism: Do Ecosystem Ecology and Landscape Ecology Clarify the Debate?" *Acta Biotheoretica*, 46:185–206.

Berkeley, George, 1710. *A Treatise Concerning the Principles of Human Knowledge*. In A. A. Luce and T. E. Jessop, editors, *The Works of George Berkeley*, volume 2. London: Thomas Nelson and Sons.

Berker, Selim, 2017. "The Unity of Grounding". *Mind*, 127:729–77.

Bernstein, Sara and Jessica M. Wilson, 2016. "Free Will and Mental Quausation". *Journal of the American Philosophical Association*, 2:310–31.

Biggs, Stephen and Jessica M. Wilson, 2017. "The A Priority of Abduction". *Philosophical Studies*, 174:735–58.

Biggs, Stephen and Jessica M. Wilson, 2019. "Abduction versus Conceiving in the Epistemology of Modality". In *New Directions in the Epistemology of Modality, Special Issue of Synthese*, Antonella Mallozzi, editor. Published online at https://doi.org/10.1007/s11229-019-02358-8.

Biggs, Stephen and Jessica M. Wilson, 2020. "Abductive Two-Dimensionalism: A New Route to the A Priori Identification of Necessary Truths". *Synthese* 197:59–93. First published online in 2017 at https://doi.org/10.1007/s11229-017-1444-6.

Bird, Alexander, 2001. "Necessarily, Salt Dissolves in Water". *Analysis*, 61:267–74.

Bird, Alexander, 2002. "On Whether Some Laws are Necessary". *Analysis*, 62:257–70.

Bird, Alexander, 2007. *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press.

Bliss, Ricki and Kelly Trogdon, 2016. "Metaphysical Grounding". *The Stanford Encyclopedia of Philosophy* (Winter 2016 edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/win2016/entries/grounding/. Accessed August 31, 2020.

Blitz, David, 1994. *Emergent Evolution: Qualitative Novelty and the Levels of Reality.* Dordrecht: Springer.

Block, Ned and Jerry A. Fodor, 1972. "What Psychological States Are Not". *Philosophical Review*, 81:159–81.

Block, Ned and Robert C. Stalnaker, 1999. "Conceptual Analysis, Dualism, and the Explanatory Gap". *Philosophical Review*, 108:1–46.

Boghossian, Paul, 1996. "Analyticity Reconsidered". *Noûs*, 30:360–91.

Bohm, David, 1952. "A Suggested Interpretation of Quantum Theory in Terms of 'Hidden Variables', Parts I and II". *The Physical Review*, 85:166–79, 180–93.

Bokulich, Alisa, 2014. "Metaphysical Indeterminacy, Properties, and Quantum Theory". *Res Philosophica*, 91:449–75.

Bonjour, Lawrence, 1998. *In Defence of Pure Reason*. Cambridge: Cambridge University Press.

Bouchard, Frederic and Philippe Huneman, editors, 2013. *From Groups to Individuals: Evolution and Emerging Individuality*. Cambridge: MIT Press.

Boyd, Richard, 1980. "Materialism without Reduction: What Physicalism does Not Entail". In Ned Block, editor, *Readings in the Philosophy of Psychology*, volume 1, 67–106. Cambridge, MA: Harvard University Press.

Broad, C. D., 1925. *Mind and Its Place in Nature*. (From the 1923 Tanner Lectures at Cambridge.) Cambridge: Kegan Paul.

Bryant, Amanda, 2020. "Physicalism". In Michael J. Raven, editor, *The Routledge Handbook of Metaphysical Grounding*, 484–500. New York: Routledge.

Byrne, Alex, 1994. *The Emergent Mind*. Ph.D. thesis, Princeton University,

Byrne, Alex, 2020. "Inverted Qualia". *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/fall2020/entries/qualia-inverted/. Accessed August 31, 2020.

Calosi, Claudio and Jessica M. Wilson, 2019. "Quantum Metaphysical Indeterminacy". *Philosophical Studies*, 176: 2599–627. First published online in 2018 at https://doi.org/10.1007/s11098-018-1143-2.

Camazine, Scott, Jean-Louis Deneubourg, Nigel R. Franks, James Sneyd, Guy Theraulaz, and Eric Bonabeau, 2018. *Self-organization in Biological Systems*. Princeton, NJ: Princeton University Press.

Campbell, Keith, 1990. *Abstract Particulars*. Oxford: Blackwell.

Carere, Claudio, Simona Montanino, Flavia Moreschini, Francesca Zoratto, Flavia Chiarotti, Daniela Santucci, and Enrico Alleva, 2009. "Aerial Flocking Patterns of Wintering Starlings, Sturnus vulgaris, under Different Predation Risk". *Animal Behaviour*, 77:101–7.

Carruth, Alexander, 2018. "Emergence, Reduction and the Identity and Individuation of Powers". *Topoi*, 39:1021–30.

Caruso, Gregg, 2012. *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*. Washington, D.C.: Lexington Books.

Chalmers, David J., 1996. *The Conscious Mind*. Oxford: Oxford University Press.

Chalmers, David J., 1999. "Materialism and the Metaphysics of Modality". *Philosophy and Phenomenological Research*, 59:473–96.

Chalmers, David J., 2002. "Does Conceivability Entail Possibility?" In Tamar Gendler and John Hawthorne, editors, University Press *Conceivability and Possibility*, 145–200. New York: Oxford University Press.

Chalmers, David J., 2003. "Consciousness and its Place in Nature". In Stephen P. Stich and Ted A. Warfield, editors, *Blackwell Guide to the Philosophy of Mind*, 102–42. Malden, MA: Blackwell.

Chalmers, David J., 2004. "The Representational Character of Experience". In Brian Leiter, editor, *The Future for Philosophy*, 153–81. New York: Oxford University Press.

Chalmers, David J., 2006*a*. "Strong and Weak Emergence". In Philip Clayton and Paul Davies, editors, *The Re-Emergence of Emergence* 244–56. New York: Oxford University Press.

Chalmers, David J., 2006*b*. "The Foundations of Two-Dimensional Semantics". In Manuel Garcia-Carpintero and Josep Macia, editors, *Two-Dimensional Semantics: Foundations and Applications*, 55–140. Oxford: Clarendon Press.

Chalmers, David J., 2009. "The Two-Dimensional Argument Against Materialism". In Brian P. McLaughlin and Sven Walter, editors, *Oxford Handbook to the Philosophy of Mind*, 1–26. Oxford: Oxford University Press.

Chalmers, David J., 2012. *Constructing the World*. Oxford University Press.

Chalmers, David J. and Frank Jackson, 2001. "Conceptual Analysis and Reductive Explanation". *The Philosophical Review*, 110:315–60.

Churchland, Patricia S., 1986. *Neurophilosophy: Toward A Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press.

Churchland, Paul M., 1981. "Eliminative Materialism and the Propositional Attitudes". *Journal of Philosophy*, 78:67–90.

Churchland, Paul M., 1984. *Matter and Consciousness*. Cambridge, MA: MIT Press.

Clapp, Lenny, 2001. "Disjunctive Properties: Multiple Realizations". *Journal of Philosophy*, 98:111–36.

Clarke, Randolph and Justin Capes, 2017. "Incompatibilist (Nondeterministic) Theories of Free Will". *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/spr2017/entries/incompatibilism-theories/. Accessed August 31, 2020.

Clarke, Randolph, 1992. "Free Will and the Conditions of Moral Responsibility". *Philosophical Studies*, 66:53–72.

Clarke, Randolph, 1999. "Nonreductive Physicalism and the Causal Powers of the Mental". *Synthese*, 51:295–322.

Clarke, Randolph, 2003. *Libertarian Accounts of Free Will*. New York: Oxford University Press.

Cohen, Stewart, 2010. "Bootstrapping, Defeasible Reasoning, and a Priori Justification". *Philosophical Perspectives*, 24:141–59.

Cohen-Tannoudji, Claude, Bernard Diu, and Franck Laloë, 1977. *Quantum Mechanics*, volume 1. Paris: John Wiley & Sons, 2nd edition.

Conee, Earl, 1994. "Phenomenal Knowledge". *Australasian Journal of Philosophy*, 72: 136–50.

Couzin, I. D., 2007. "Collective Minds". *Nature*, 445:4.

Crane, Tim, 2001. "The Significance of Emergence". In Carl Gillett and Barry M. Loewer, editors, *Physicalism and Its Discontents*, 207–24. Cambridge: Cambridge University Press.

Crane, Tim and D. H. Mellor, 1990. "There Is No Question of Physicalism". *Mind*, 99: 185–206.

Craver, Carl F., 2001. "Mechanisms, and Hierarchy". *Philosophy of Science*, 68:53–74.

Cucu, Alin C. and J. Brian Pitts, 2019. "How Dualists Should (Not) Respond to the Objection From Energy Conservation". *Mind and Matter*, 17:95–121.

Cummins, Robert, 1975. "Functional Analysis". *Journal of Philosophy*, 72:741–64.

Cummins, Robert, 1983. "Analysis and Subsumption in the Behaviorism of Hull", Philosophy of Science, 50:96–111.

Cunningham, Bryon, 2001. "The Reemergence of Emergence". *Philosophy of Science*, 68:S62–S75. Supplement: Proceedings of the 2000 Biennial Meeting of the Philosophy of Science Association.

Damper, R. I., 2000. "Editorial for the Special Issue on 'Emergent Properties of Complex Systems' ". *International Journal of Systems Science*, 31:811–18.

Dasgupta, Shamik, 2014. "The Possibility of Physicalism". *Journal of Philosophy*, 111:557–92.

Davidson, Donald, 1963. "Actions, Reasons, and Causes". *Journal of Philosophy*, 60:685–700.

Davidson, Donald, 1970. "Mental Events". In L. Foster and J. Swanson, editors, *Experience and Theory*, 207–224. Amherst, MA: Massachusetts University Press. Reprinted in Davidson 1980/2001.

Davidson, Donald, 1980/2001. *Essays on Actions and Events*. Oxford: Oxford University Press.

deRosset, Louis, 2011. "What Is the Grounding Problem?" *Philosophical Studies*, 156: 173–97.

Descartes, René, 1641–7/1984. "Meditations". In Robert Stoothoff, John Cottingham, and Dugald Murdoch, editors, *The Philosophical Writings of Descartes*, volume II. Cambridge: Cambridge University Press.

Dosanjh, Ranpal, 2014. *A Defense of Reductive Physicalism*. Ph.D. thesis, University of Toronto.

Dosanjh, Ranpal, 2019. "Token-Distinctness and the Disjunctive Strategy". *Erkenntnis*, 1–18. Published online at https://doi.org/10.1007/s10670-019-00127-0.

Douven, Igor, 2017. "Abduction". *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/sum2017/entries/abduction/. Accessed August 31, 2020.

Dowell, Janice, 2006. "Formulating the Thesis of Physicalism: An Introduction". *Philosophical Studies*, 131:1–23.

Dretske, Fred I., 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Dummett, Michael, 1975. "Wang's Paradox". *Synthese*, 30:201–32.

Earman, John, 1976. "Causation: A Matter of Life and Death". *The Journal of Philosophy*, 73:5–25.

Ehring, Douglas, 1996. "Mental Causation, Determinables, and Property Instances". *Noûs*, 30:461–80.

Ehring, Douglas, 2003. "Part-whole Physicalism and Mental Causation". *Synthese*, 136:359–88.

Ekstrom, Laura W., 2001. *Agency and Responsibility: Essays on the Metaphysics of Freedom*. Boulder, CO: Westview Press.

Emmeche, Claus, Simo Koppe, and Frederick Stjernfelt, 1997. "Explaining Emergence: Toward an Ontology of Levels". *Journal for General Philosophy of Science/Zeitschrift für Allgemeine Wissenschaftstheorie*, 28:83–119.

Evans, Gareth, 1983. "Can There Be Vague Objects?" *Analysis*, 38:208.

Ewing, Alfred C., 1951. *The Fundamental Questions of Philosophy*. Oxford: Routledge.

Fales, Evan, 1990. *Causation and Universals*. New York: Routledge.

Farrell, B. A., 1950. "Experience". *Mind*, 59:170–98.

Feigl, Herbert, 1959. "The 'Mental' and the 'Physical' ". In Grover Maxwell, Herbert Feigl, and Michael Scriven, editors, *Minnesota Studies in the Philosophy of Science*, volume 2. Minneapolis, MN: University of Minnesota Press.

Feigl, Herbert, 1967. *The "Mental" and the "Physical": The Essay and a Postscript*. Minneapolis, MN: University of Minnesota Press.

Feyerabend, Paul K., 1963. "Mental Events and the Brain". *Journal of Philosophy*, 40:295–6.

Feynman, Richard, 1963. *The Feynman Lectures on Physics*, Volumes I-III. Boston, MA: Addison Wesley.

Feynman, Richard, 1965. *The Character of Physical Law*. Cambridge, MA: MIT Press.

Field, Hartry, 2003. "Causation in a Physical World". In Michael Loux and Dean W. Zimmerman, editors, *The Oxford Handbook of Metaphysics*, 435–460. Oxford: Oxford University Press.

Fine, Kit, 2001. "The Question of Realism". *Philosophers' Imprint*, 1:1–30.

Fine, Kit, 2003. "The Non-Identity of a Material Thing and its Matter". *Mind*, 112:195–234.

Flanagan, Owen, 1992. *Consciousness Reconsidered*. Cambridge, MA: MIT Press.

Fodor, Jerry, 1974. "Special Sciences (Or, The Disunity of Science as a Working Hypothesis)". *Synthese*, 28:77–115.

Forster, Malcolm and Alexey Kryukov, 2003. "The Emergence of the Macroworld: A Study of Intertheory Relations in Classical and Quantum Mechanics". *Philosophy of Science*, 70:1039–51.

Francescotti, Robert, 2007. "Emergence". *Erkenntnis*, 67:47–63.

Frankfurt, Harry G., 1971. "Freedom of the Will and the Concept of a Person". *The Journal of Philosophy*, 68:5–20.

Franklin, Christopher Evan, 2018. *A Minimal Libertarianism: Free Will and the Promise of Reduction*. New York: Oxford University Press.

Frege, Gottlob, 1892/2010. "On Sense and Reference". In Darragh Byrne and Max Kölbel, editors, *Arguing About Language*, 36–56. London: Routledge.

Frigg, Roman, 2009. "GRW Theory: Ghirardi, Rimini, Weber Model of Quantum Mechanics". In Klaus Hentschel Daniel Greenberger, Brigitte Falkenburg, and Friedel Weinert, editors, *Compendium of Quantum Physics: Concepts, Experiments, History and Philosophy*, 266–70. Berlin: Springer.

Funkhouser, Eric, 2006. "The Determinable-Determinate Relation". *Noûs*, 40:548–69.

Garcia, Robert K., 2014. "Closing in on Causal Closure". *Journal of Consciousness Studies*, 21:96–109.

Gertler, Brie, 2002. "Explanatory Reduction, Conceptual Analysis, and Conceivability Arguments About the Mind". *Noûs*, 36:22–49.

Gertler, Brie, 2006. "Consciousness and Qualia Cannot Be Reduced". In Robert J. Stainton, editor, *Contemporary Debates in Cognitive Science*, 202–16. Malden, MA: Blackwell.

Ghirardi, G., A. Rimini, and T. Weber, 1986. "Unified Dynamics for Microscopic and Macroscopic Systems". *The Physical Review* D, 34:470–9.

Gibb, Sophie C., Robin Findlay Hendry, and Tom Lancaster, 2018. *The Routledge Handbook of Emergence*. Oxford: Routledge.

Gibb, Sophie C., 2013. "The Entailment Problem and the Subset Account of Property Realization". *Australasian Journal of Philosophy*, 92:551–66.

Gibbard, Allan, 1975. "Contingent Identity". *Journal of Philosophical Logic*, 4:187–222.

Gillett, Carl, 2002*a*. "The Dimensions of Realization: A Critique of the Standard View". *Analysis*, 62:316–23.

Gillett, Carl, 2002*b*. "The Varieties of Emergence: Their Purposes, Obligations and Importance". *Grazer Philosophische Studien*, 65:95–121.

Gillett, Carl, 2010. "Moving Beyond the Subset Model of Realization: The Problem of Qualitative Distinctness in the Metaphysics of Science". *Synthese*, 177:165–92.

Gillett, Carl, 2016. *Reduction and Emergence in Science and Philosophy*. Cambridge: Cambridge University Press.

Gillett, Carl and Barry M. Loewer, editors, 2001. *Physicalism and Its Discontents*. Cambridge: Cambridge University Press.

Ginet, Carl, 1990. *On Action*. Cambridge: Cambridge University Press.

Ginet, Carl, 2002. "Reasons Explanations of Action: Causalist Versus Noncausalist Accounts". In Robert H. Kane, editor, *The Oxford Handbook on Free Will*, 386–405. Oxford: Oxford University Press.

Grandy, Richard E., 2016. "Sortals". *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/win2016/entries/sortals/. Accessed August 31, 2020.

Grandy, Richard E., 2007. "Artifacts: Parts and Principles". In Eric Margolis and Stephen Laurence, editors, *Creations of the Mind: Theories of Artifacts and Their Representation*, 18–32. New York: Oxford University Press.

Greiner, Walter, 1996. "The Discovery of the Weak Interaction". In Walter Greiner and Berndt Müller, editors, *Gauge Theory of Weak Interactions*, 1–24. Berlin, Heidelberg: Springer.

Guay, Alexandre and Olivier Sartenaer, 2016. "A New Look at Emergence. Or When After is Different". *European Journal for Philosophy of Science*, 6:297–322.

Haggard, Patrick and Benjamin W. Libet, 2001. "Conscious Intention and Brain Activity". *Journal of Consciousness Studies*, 8:47–63.

Hall, Ned, 2004. "Two Concepts of Causation". In John Collins, Ned Hall, and Laurie Paul, editors, *Causation and Counterfactuals*, 225–76. Stanford, MA: MIT Press.

Harman, Gilbert H., 1965. "The Inference to the Best Explanation". *Philosophical Review*, 74:88–95.

Harman, Gilbert H., 1990. "The Intrinsic Quality of Experience". In James Tomberlin, editor, *Action Theory and the Philosophy of Mind*, volume 4 of *Philosophical Perspectives*, 31–52. Atascadero, CA: Ridgeview. Reprinted in Harman 1999.

Harman, Gilbert H., 1999. *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press.

Haug, Matthew C., 2010. "Realization, Determination, and Mechanisms". *Philosophical Studies*, 150:313–30.

Hawthorne, John, 2001. "Causal Structuralism". *Noûs*, 35:361–78.

Hawthorne, John, 2002. "Deeply Contingent a Priori Knowledge". *Philosophy and Phenomenological Research*, 65:247–69.

Hawthorne, John and Philip Pettit, 1996. "Strategies for Free Will Compatibilists". *Analysis*, 56.4:191–201. Hawthorne originally published this article under the name 'John O'Leary-Hawthorne'.

Heil, John, 1992. *The Nature of True Minds*. Cambridge: Cambridge University Press.

Heil, John, 2003*a*. *From an Ontological Point of View*. Oxford: Clarendon Press.

Heil, John, 2003*b*. "Levels of Reality". *Ratio*, 16:205–21.

Hellie, Benj, 2007. " 'There's Something It's Like' and the Structure of Consciousness". *Philosophical Review*, 116:441–63.

Hellie, Benj, 2019*a*. "An Analytic-Hermeneutic History of Consciousness". In Kelly Michael Becker and Iain Thompson, editors, *Cambridge History of Philosophy: 1945-2015*, 74–89. Cambridge: Cambridge University Press.

Hellie, Benj, 2019*b*. "Semantic Gaps and Protosemantics". In Acacio de Barros and Carlos Montemayor, editors, *Mind and Quanta*, 201–22. Berlin: Springer.

Hellie, Benj, forthcoming. *Out of This World: Logical Mentalism and the Philosophy of Mind*. New York: Oxford University Press.

Hellman, Geoffrey and Frank Thompson, 1975. "Physicalism: Ontology, Determination, and Reduction". *The Journal of Philosophy*, 72:551–64.

Hempel, Carl G., 1979. "Comment at a Symposium on Nelson Goodman's *Ways of Worldmaking*". *Presented at the 76th Annual Meeting of the American Philosophical Association*.

Hempel, Carl G. and Paul Oppenheim, 1948. "Studies in the Logic of Explanation". *Philosophy of Science*, 15:135–75.

Hendry, Robin Findlay, 2010. "Emergence vs. Reduction in Chemistry". In Cynthia F. Macdonald and Graham Macdonald, editors, *Emergence in Mind*, 205–21. New York: Oxford University Press.

Hendry, Robin Findlay, 2017. "Prospects for Strong Emergence in Chemistry". In Michele Paolini Paoletti and Francesco Orilia, editors, *Philosophical and Scientific Perspectives on Downward Causation*, 146–63. New York: Routledge.

Hill, Christopher S., 2009. *Consciousness*. New York: Cambridge University Press.

Hitchcock, Christopher Read, 1996. "The Role of Contrast in Causal and Explanatory Claims". *Synthese*, 107:395–419.

Hodgeson, Shadworth, 1962. *The Theory of Practice: An Ethical Enquiry*. London: Longmans, Green, Reader, and Dyer.

Hooker, C. A., 2004. "Asymptotics, Reduction and Emergence". *British Journal for the Philosophy of Science*, 55:435–79.

Horgan, Terence, 1982. "Supervenience and Microphysics". *Pacific Philosophical Quarterly*, 63:29–43.

Horgan, Terence, 1989. "Mental Quausation". *Philosophical Perspectives 3: Philosophy of Mind and Action Theory*, 47–76.

Horgan, Terence, 1993. "From Supervenience to Superdupervenience: Meeting the Demands of a Material World". *Mind*, 102:555–86.

Howell, Robert J., 2009. "Emergentism and Supervenience Physicalism". *Australasian Journal of Philosophy*, 87:83–98.

Huggett, Nick and Christian Wüthrich, 2013. "Emergent Spacetime and Empirical (in)Coherence". *Studies in History and Philosophy of Modern Physics*, 44:276–85.

Humphreys, Paul, 1996. "Aspects of Emergence". *Philosophical Topics*, 24:53–70.

Humphreys, Paul, 1997. "How Properties Emerge". *Philosophy of Science*, 64:1–17.

Humphreys, Paul, 2016. *Emergence: A Philosophical Account*. Oxford: Oxford University Press.

Huxley, Thomas, 1874. "On the Hypothesis That Animals Are Automata, and its History". *Fortnightly Review*, 95:555–80.

Jackson, Frank, 1982. "Epiphenomenal Qualia". *Philosophical Quarterly*, 32:127–36.

Jackson, Frank, 1986. "What Mary Didn't Know". *The Journal of Philosophy*, 83:291–5.

Jackson, Frank, 1994. "Armchair Metaphysics". In John O'Leary-Hawthorne and Michaelis Michael, editors, *Philosophy in Mind*, 23–42. Dordrecht: Kluwer.

Jaworski, William, 2002. "Multiple-Realizability, Explanation and the Disjunctive Move". *Philosophical Studies*, 108:289–308.

Jenkins, C. S., 2013. “Explanation and Fundamentality”. In Miguel Hoeltje, Benjamin Schnieder, and Alex Steinberg, editors, *Varieties of Dependence (Basic Philosophical Concepts Series)*, 211–42. Munich: Philosophia Verlag.

Judd, D. and G. Wyszecki, 1975. *Color in Business, Science, and Industry*. New York: Wiley.

Kane, Gordon, 1993. *Modern Elementary Particle Physics: the Fundamental Particles and Forces*. Boulder, CO: Westview Press.

Kauffman, Stuart, 1993. *The Origins of Order*. Oxford: Oxford University Press.

Kauffman, Stuart, 1995. *At Home in the Universe: The Search for Laws of Self-organization and Complexity*. New York: Oxford University Press.

Kim, Jaegwon, 1984. “Concepts of Supervenience”. *Philosophy and Phenomenological Research*, 45:153–76.

Kim, Jaegwon, 1989. “The Myth of Nonreductive Materialism”. *Proceedings and Addresses of the American Philosophical Association*, 63:31–47. Reprinted in Kim 1993*b*.

Kim, Jaegwon, 1990. “Supervenience as a Philosophical Concept”. *Metaphilosophy*, 21:1–27. Reprinted in Kim 1993*b*.

Kim, Jaegwon, 1992*a*. “ “Downward Causation” in Emergentism and Nonreductive Physicalism”. In Ansgar Beckermann, Hans Flohr, and Jaegwon Kim, editors, *Emergence or Reduction?: Prospects for Nonreductive Physicalism*, 119–38. Berlin: De Gruyter.

Kim, Jaegwon, 1992*b*. “Multiple Realization and the Metaphysics of Reduction”. *Philosophy and Phenomenological Research*, 52:1–26. Reprinted in Kim 1993*b*.

Kim, Jaegwon, 1993*a*. “The Non-Reductivist’s Troubles with Mental Causation”. In John Heil and Alfred Mele, editors, *Mental Causation*, 189–210. Oxford: Oxford University Press. Reprinted in Kim 1993*b*.

Kim, Jaegwon, 1993*b*. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.

Kim, Jaegwon, 1998. *Mind in a Physical World*. Cambridge, MA: MIT Press.

Kim, Jaegwon, 1999. “Making Sense of Emergence”. *Philosophical Studies*, 95:3–36.

Kim, Jaegwon, 2006. “Emergence: Core ideas and issues”. *Synthese*, 151:547–59.

Kim, Jaegwon, 2010. “Thoughts on Sydney Shoemaker’s Physical Realization”. *Philosophical Studies*, 148:101–12.

Kim, Jaegwon, 2015. “What Could Pair a Nonphysical Soul to a Physical Body?” In Keith Augustine and Michael Martin, editors, *The Myth of an Afterlife: The Case against Life After Death*, 335–47. Lanham, MD: Rowman & Littlefield.

Kirchhoff, Michael, 2014. “In Search of Ontological Emergence: Diachronic, But Non-Supervenient”. *Axiomathes*, 24:89–116.

Kirk, Robert, 2019. “Zombies”. *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/spr2019/entries/zombies/. Accessed August 31, 2020.

Klee, Robert, 1984. “Micro-Determinism and Concepts of Emergence”. *Philosophy of Science*, 51:44–63.

Kleman, Maurice and Oleg Lavrentovich, 2003. *Soft Matter Physics: An Introduction*. New York: Springer.

Knox, Eleanor, 2013. “Effective Spacetime Geometry”. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44:346–56.

Knox, Eleanor, 2018. “Physical Relativity From a Functionalist Perspective”. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 67:118–24.

Korman, Daniel Z., 2020. "Ordinary Objects". *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/fall2020/entries/ordinary-objects/. Accessed August 31, 2020.

Koslicki, Kathrin, 2008. *The Structure of Objects*. New York: Oxford University Press.

Koslicki, Kathrin, 2012. "Varieties of Ontological Dependence". In Fabrice Correia and Benjamin Schnieder, editors, *Metaphysical Grounding: Understanding the Structure of Reality*, 186. Cambridge: Cambridge University Press.

Koslicki, Kathrin, 2016. "Where Grounding and Causation Part Ways: Comments on Jonathan Schaffer". *Philosophical Studies*, 173:101–12.

Kovacs, David, 2019. "The Myth of the Myth of Supervenience". *Philosophical Studies*, 176:1967–1989.

Kripke, Saul, 1972/80. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Ladyman, James, James Lambert, and Karoline Wiesner, 2013. "What is a Complex System?" *European Journal for Philosophy of Science*, 3:33–67.

Ladyman, James and Don Ross, 2007. *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.

Lam, Vincent and Christian Wüthrich, 2018. "Spacetime is as Spacetime Does". *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 64:39–51.

Lamb, Maurice, 2015. *Characteristics of Non-reductive Explanations in Complex Dynamical Systems Research*. Ph.D. thesis, University of Cincinnati.

Lancaster, Tom and Mark Pexton, 2015. "Reduction and Emergence in the Fractional Quantum Hall State". *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 52:343–57.

Langsam, Harold, 2011. *The Wonder of Consciousness: Understanding the Mind Through Philosophical Reflection*. Cambridge, MA: MIT Press.

Leibniz, Gottfried Wilhelm, 1714. *The Monadology*.

Lepore, Ernest and Barry M. Loewer, 1987. "Mind Matters". *The Journal of Philosophy*, 84:630–42.

Lepore, Ernest and Barry M. Loewer, 1989. "More on Making Mind Matter". *Philosophical Topics*, 17:175–91.

Levin, Janet, 2016. "Review of Thomas Polger and Stuart Shapiro, *The Multiple Realization Book*". *Notre Dame Philosophical Reviews*. Published online at https://ndpr.nd.edu/news/the-multiple-realizat.

Levin, Simon A., 2005. "Self-organization and the Emergence of Complexity in Ecological Systems". *BioScience*, 55:1075–9.

Levine, Joseph M., 1983. "Materialism and Qualia: The Explanatory Gap". *Pacific Philosophical Quarterly*, 64:354–61.

Levine, Joseph M., 2001. *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press.

Lewes, G. H., 1875. *Problems of Life and Mind*. London: Kegan Paul, Trench, Trübner & Co.

Lewis, David, 1966. "An Argument for the Identity Theory". *The Journal of Philosophy*, 63:17–25. Reprinted in Lewis 1983*b*.

Lewis, David, 1981. "Are We Free to Break the Laws?" *Theoria*, 47:113–21.

Lewis, David, 1983*a*. "New Work for a Theory of Universals". *Australasian Journal of Philosophy*, 61:343–77.

Lewis, David, 1983*b*. *Philosophical Papers*, volume I. Oxford: Oxford University Press.

Lewis, David, 1986. *On the Plurality of Worlds*. London: Blackwell.

Lewis, David, 1990. "What Experience Teaches". In William G. Lycan, editor, *Mind and Cognition: An Anthology*, 29–57. Oxford: Blackwell. Reprinted in Lewis 1999.

Lewis, David, 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Lewtas, Patrick, 2013. "Emergence and Consciousness". *Philosophy*, 88:527–53.

Libet, Benjamin W., 2002. "The Timing of Mental Events: Libet's Experimental Findings and Their Implications". *Consciousness and Cognition*, 11:291–9.

Libet, Benjamin W., 1999. "Do We Have Free Will?" *Journal of Consciousness Studies*, 6: 47–57.

List, Christian and Peter Menzies, 2009. "Nonreductive Physicalism and the Limits of the Exclusion Principle". *Journal of Philosophy*, 106:475–502.

Locke, John, 1690. *An Essay Concerning Human Understanding*.

Loewer, Barry M., 2001. "From Physics to Physicalism". In Carl Gillett and Barry M. Loewer, editors, *Physicalism and Its Discontents*, 37–56. Cambridge: Cambridge University Press.

Lowe, E. J., 1989. "What Is a Criterion of Identity?" *Philosophical Quarterly*, 39:1–21.

Lowe, E. J., 1998. *The Possibility of Metaphysics: Substance, Identity, and Time*. New York: Oxford University Press.

Lowe, E. J., 2007. "Sortals and the Individuation of Objects". *Mind and Language*, 22:514–33.

MacBride, Fraser, 1999. "Could Armstrong have been a Universal?" *Mind*, 108:471–501.

Macdonald, Cynthia and Graham Macdonald 1986. "Mental Causes and Explanation of Action". In L. Stevenson, R. Squires, and J. Haldane, editors, *Mind, Causation, and Action*, 145–58. Oxford: Basil Blackwell.

Macdonald, Cynthia and Graham Macdonald, 1995. "How to be Psychologically Relevant". In *Philosophy of Psychology: Debates on Psychological Explanation*, 60–77. Oxford: Blackwell.

Maddy, Penelope, 1997. *Naturalism in Mathematics*. New York: Oxford University Press.

Malcolm, Norman, 1968. "The Conceivability of Mechanism". *The Philosophical Review*, 77:45–72.

Margolis, Eric and Stephen Laurence, editors, 2007. *Creations of the Mind: Theories of Artifacts and Their Representation*. Oxford: Oxford University Press.

Martin, C. B., 1996. "Properties and Dispositions". In Tim Crane, editor, *Dispositions: A Debate*, 71–81. London: Routledge.

Mayoral, E. and A. Robledo, 1970. "A Recent Appreciation of the Singular Dynamics at the Edge of Chaos". In Marcel Ausloos and Michel Dirickx, editors, *The Logistic Map and the Route to Chaos: From the Beginnings to Modern Applications*, 339–54. Berlin: Springer.

McCann, Hugh, 1998. *The Works of Agency: On Human Action, Will, and Freedom*. Ithaca, NY: Cornell University Press.

McDaniel, Kris, 2001. "Tropes and Ordinary Physical Objects". *Philosophical Studies*, 104:269–90.

McKenna, Michael and D. Justin Coates, 2020. "Compatibilism". In *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/sum2020/entries/compatibilism/. Accessed August 31, 2020.

McLaughlin, Brian P., 1992. "The Rise and Fall of British Emergentism". In Ansgar Beckerman, Hans Flohr, and Jaegwon Kim, editors, *Emergence or Reduction? Essays on the Prospects of Non-reductive Physicalism*, 49–93. Berlin: De Gruyter.

McLaughlin, Brian P. and Karen Bennett, 2018. "Supervenience". In *The Stanford Encyclopedia of Philosophy* (Winter 2018 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/win2018/entries/supervenience/. Accessed August 31, 2020.

McLaughlin, Brian P., 2007. "Mental Causation and Shoemaker-Realization". *Erkenntnis*, 67:149–72.

McLeish, Tom C. B., Mark Pexton, and Tom Lancaster, 2019. "Emergence and Topological Order in Classical and Quantum Systems". *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 66:155–69.

McLeish, Tom C. B., 2017. "Strong Emergence and Downward Causation in Biological Physics". *Philosophica*, 92:113–38.

Megill, Jason, 2013. "A Defense of Emergence". *Axiomathes*, 23:597–615.

Mele, Alfred, 2009. *Effective Intentions: The Power of Conscious Will*. Oxford: Oxford University Press.

Melnyk, Andrew, 1997. "How to Keep the 'Physical' in Physicalism". *The Journal of Philosophy*, 94:622–37.

Melnyk, Andrew, 1999. "Supercalifragilisticexpialidocious". *Noûs*, 33:144–54.

Melnyk, Andrew, 2003. *A Physicalist Manifesto: Thoroughly Modern Materialism*. New York: Cambridge University Press.

Melnyk, Andrew, 2006. "Realization-based Formulations of Physicalism". *Philosophical Studies*, 131:127–55.

Menon, Tarun and Craig Callender, 2013. "Ch-Ch-Changes Philosophical Questions Raised by Phase Transitions". In Robert W. Batterman, editor, *The Oxford Handbook of Philosophy of Physics*, 189–212. New York: Oxford University Press.

Merricks, Trenton, 2003. *Objects and Persons*. Oxford: Clarendon Press.

Messiah, Albert, 1970. *Quantum Mechanics*. Amsterdam: North-Holland.

Mill, John S., 1843/1973. *A System of Logic*. Toronto: University of Toronto Press.

Mitchell, Sandra D., 2012. "Emergence: Logical, Functional and Dynamical". *Synthese*, 185:171–86.

Montero, Barbara, 2003. "Varieties of Causal Closure". In Sven Walter and Heinz-Dieter Heckmann, editors, *Physicalism and Mental Causation*, 173–87. Exeter: Imprint Academic.

Montero, Barbara, 2006. "Physicalism in an Infinitely Decomposable World". *Erkenntnis*, 64:177–91.

Monton, Bradley, 2004. "The Problem of Ontology on Spontaneous Collapse Theories". *Studies in History and Philosophy of Modern Physics*, 35:407–21.

Moore, G. E., 1903. *Principia Ethica*. Cambridge: Cambridge University Press.

Mørch, Hedda Hassel, 2018. "The Evolutionary Argument for Phenomenal Powers". *Philosophical Perspectives*, 31: 293–316.

Mørch, Hedda Hassel (2020). "The Phenomenal Powers View and the Meta-Problem of Consciousness". *Journal of Consciousness Studies*, 27:131–42.

Morgan, C. Lloyd, 1923. *Emergent Evolution*. London: Williams & Norgate.

Morris, Kevin, 2011. "Subset Realization, Parthood, and Causal Overdetermination". *Pacific Philosophical Quarterly*, 92:363–79.

Morris, Kevin, 2013. "On Two Arguments for Subset Inheritance". *Philosophical Studies*, 163:197–211.

Morris, Kevin, 2014. "Supervenience Physicalism, Emergentism, and the Polluted Supervenience Base". *Erkenntnis*, 79:351–65.

Morris, Kevin, 2018. *Physicalism Deconstructed: Levels of Reality and the Mind–Body Problem*. Cambridge: Cambridge University Press.

Morrison, Margaret, 2006. "Emergence, Reduction, and Theoretical Principles: Rethinking Fundamentalism". *Philosophy of Science*, 73:876–87.

Morrison, Margaret, 2012. "Emergent Physics and Micro-Ontology". *Philosophy of Science*, 79:141–66.

Mossio, Matteo, Leonardo Bich, and Alvaro Moreno, 2013. "Emergence, Closure and Inter-Level Causation in Biological Systems". *Erkenntnis*, 78:153–78.

Nagel, Ernest, 1961. *The Structure of Science*. London: Routledge & Kegan Paul.

Nagel, Thomas, 1974. "What is it Like to be a Bat?" *The Philosophical Review*, 83:435–50.

Nagel, Thomas, 1979. "Panpsychism". In Thomas Nagel, editor, *Mortal Questions*, 181–95. Binghamton, NY: Cambridge University Press.

Nemirow, Laürence, 2006. "So This is What It's Like: A Defense of the Ability Hypothesis". In Torin Alter and Sven Walter, editors, *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, 32–51. New York: Oxford University Press.

Newman, David, 1996. "Emergence and Strange Attractors". *Philosophy of Science*, 63: 245–61.

Ney, Alyssa, 2008. "Defining Physicalism". *Philosophy Compass*, 3:1033–48.

Ney, Alyssa, 2010. "Convergence on the Problem of Mental Causation: Shoemaker's Strategy for (Nonreductive?) Physicalists". *Philosophical Issues*, 20:438–45.

Nida-Rümelin, Martine and Donnchadh O Conaill, 2019. "Qualia: The Knowledge Argument". *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/win2019/entries/qualia-knowledge/. Accessed August 31, 2020.

Noordhof, Paul, 2010. "Emergent Causation and Property Causation". In Cynthia Macdonald and Graham F. Macdonald, editors, *Emergence in Mind*, 69–99. Oxford: Oxford University Press.

Nozick, Robert, 1995. "Choice and Indeterminism". In Timothy O'Connor, editor, *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, 101–14. New York: Oxford University Press.

O'Connor, Timothy, 1994. "Emergent properties". *American Philosophical Quarterly*, 31: 91–104.

O'Connor, Timothy, 2000. *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.

O'Connor, Timothy, 2005. "Free Will". In John Martin Fischer, editor, *Free Will: Critical Concepts in Philosophy*, 7–32. New York: Routledge. Originally published in *The Stanford Encyclopedia of Philosophy (2002 Edition)*.

O'Connor, Timothy, 2009*a*. "Agent-Causal Power". In Toby Handfield, editor, *Dispositions and Causes*, 189–214. New York: Oxford University Press.

O'Connor, Timothy, 2009*b*. "Conscious Willing and the Emerging Sciences of Brain and Behavior". In Nancey Murphy, George Ellis, and Timothy O'Connor, editors, *Downward Causation and the Neurobiology of Free Will*, 173–86. Berlin: Springer.

O'Connor, Timothy and Hong Yu Wong, 2005. "The Metaphysics of Emergence". *Noûs*, 39:658–78.

O'Connor, Timothy and Hong Yu Wong, 2015. "Emergent Properties", The Stanford Encyclopedia of Philosophy (Fall 2020 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/fall2020/entries/properties-emergent/. Accessed August 31, 2020.

Owens, David, 1989. "Levels of Explanation". *Mind*, 98:59–79.

Page, R. E. and Sandra D. Mitchell, 1991. "Self Organization and Adaptation in Insect Societies". *Proceedings of the Philosophy of Science Association*, 2:289–98.

Page, R. E. and Sandra D. Mitchell, 1998. "Self Organization and the Evolution of Division of Labor". *Apidologie*, 29:101–20.

Paolini Paoletti, Michele, 2017. *The Quest for Emergence*. Munich: Philosophia.
Paolini Paoletti, Michele, 2018*a*. "Emergent Powers". *Topoi*, 1–14.
Paolini Paoletti, Michele, 2018*b*. "Formulating Emergence". *Ratio*, 31:1–18.
Papineau, David, 1993. *Philosophical Naturalism*. Oxford: Basil Blackwell.
Papineau, David, 2001. "The Rise of Physicalism". In Carl Gillett and Barry M. Loewer, editors, *Physicalism and Its Discontents*, 3–36. Cambridge: Cambridge University Press.
Paul, L. A., 2002. "Logical Parts". *Noûs*, 36:578–96.
Paul, L. A., 2010. "The Puzzles of Material Constitution". *Philosophy Compass*, 5:579–90.
Paull, C. P. and Theodore Sider, 1992. "In Defense of Global Supervenience". *Philosophy and Phenomenological Research*, 32:830–45.
Peacocke, Christopher, 1993. "How Are A Priori Truths Possible?" *European Journal of Philosophy*, 1:175–99.
Peacocke, Christopher, 2003. *The Realm of Reason*. New York: Oxford University Press.
Pearson, Olley, 2018. "Emergence, Dependence, and Fundamentality". *Erkenntnis*, 83: 391–402.
Pereboom, Derk, 2002. "Robust Non-reductive Materialism". *Journal of Philosophy*, 99: 499–531.
Pereboom, Derk, 2011. *Consciousness and the Prospects of Physicalism*. New York: Oxford University Press.
Pereboom, Derk and Hilary Kornblith, 1991. "The Metaphysics of Irreducibility". *Philosophical Studies*, 63:125–45.
Perry, John, 2001. *Knowledge, Possibility, and Consciousness*. Cambridge, MA: MIT Press, second edition.
Pettit, Philip, 1995. "Microphysicalism, Dottism, and Reduction". *Analysis*, 55:141–6.
Pitts, J. Brian, 2020. "Conservation Laws and the Philosophy of Mind: Opening the Black Box, Finding a Mirror". *Philosophia*, 48:673–707.
Poland, Jeffrey, 1994. *Physicalism; the Philosophical Foundations*. Oxford: Clarendon Press.
Polger, Thomas W., 2007. "Realization and the Metaphysics of Mind". *Australasian Journal of Philosophy*, 85:233–59.
Polger, Thomas W. and Lawrence A. Shapiro, 2016. *The Multiple Realization Book*. Oxford: Oxford University Press.
Popper, Karl R. and John C. Eccles, 1977. *The Self and Its Brain: An Argument for Interactionism*. Oxfordshire: Springer.
Putnam, Hilary, 1967. "Psychological Predicates". In *Art, Mind, and Religion*, 37–48. Pittsburgh, PA: University of Pittsburgh Press.
Raffman, Diana, 1994. "Vagueness Without Paradox". *Philosophical Review*, 103:41–74.
Raven, Michael J., 2015. "Ground". *Philosophy Compass*, 10:322–33.
Raven, Michael J., 2016. "Fundamentality Without Foundations". *Philosophy and Phenomenological Research*, 93:607–26.
Ravenscroft, Ian, 1997. "Physical Properties". *Southern Journal of Philosophy*, 35:419–31.
Robb, David, 1997. "The Properties of Mental Causation". *Philosophical Quarterly*, 47: 178–94.
Robinson, William, 2019. "Epiphenomenalism". *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/sum2019/entries/epiphenomenalism/. Accessed August 31, 2020.
Rosen, Gideon, 2010. "Metaphysical Dependence: Grounding and Reduction". In B. Hale and A. Hoffmann, editors, *Modality: Metaphysics, Logic, and Epistemology*, 109–36. Oxford: Oxford University Press.

Rosen, J., 2007 (April 22). “Flight Patterns”. *New York Times*, 255–73.
Rosenberg, Alexander, 1994. *Instrumental Biology, or, the Disunity of Science*. Chicago, IL: University of Chicago Press.
Rueger, Alexander, 2000. “Physical Emergence, Diachronic and Synchronic”. *Synthese*, 124:297–322.
Rueger, Alexander and Patrick McGivern, 2010. “Hierarchies and Levels of Reality”. *Synthese*, 176:379–97.
Russell, Bertrand, 1912. “On the Notion of Cause”. *Proceedings of the Aristotelian Society*, 13:1–26.
Russell, Bruce, 2020. “A Priori Justification and Knowledge”. The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/sum2020/entries/apriori/. Accessed August 31, 2020.
Salmon, Nathan, 1989. “The Logic of What Might Have Been”. *Philosophical Review*, 98: 3–34.
Santos, Gil C., 2020. “Integrated-Structure Emergence and Its Mechanistic Explanation”. *Synthese*. https://doi.org/10.1007/s11229-020-02594-3.
Santos, Gil C., 2015. “Upward and Downward Causation From a Relational-Horizontal Ontological Perspective”. *Axiomathes*, 25:23–40.
Schaffer, Jonathan, 2003. “Is There a Fundamental Level?” *Noûs*, 37:498–517.
Schaffer, Jonathan, 2004. “Quiddistic Knowledge”. In Frank Jackson and Graham Priest, editors, *Lewisian Themes*, 210–30. Oxford: Oxford University Press.
Schaffer, Jonathan, 2009. “On What Grounds What”. In David J. Chalmers, David Manley, and Ryan Wasserman, editors, *Metametaphysics: New Essays on the Foundations of Ontology*, 347–83. Oxford: Oxford University Press.
Schaffer, Jonathan, 2016*a*. “Ground Rules: Lessons from Wilson”. In Ken Aizawa and Carl Gillet, editors, *Scientific Composition and Metaphysical Ground*, 143–69. London: Palgrave-Macmillan.
Schaffer, Jonathan, 2016*b*. “Grounding in the Image of Causation”. *Philosophical Studies*, 173:49–100.
Schroder, Jurgen, 1998. “Emergence: Non-Deducibility or Downward Causation?” *The Philosophical Quarterly*, 48:433–52.
Seager, William, 1999/2016. *Theories of Consciousness: An Introduction and Assessment*. Abingdon: Routledge.
Seager, William, 2012. “Emergentist Panpsychism”. *Journal of Consciousness Studies*, 19: 9–10.
Searle, John R., 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
Searle, John R., 2007. “Social Ontology and the Philosophy of Society”. In Eric Margolis and Stephen Laurence, editors, *Creations of the MInd: Theories of Artifacts and Their Representation*, 3–17. New York: Oxford University Press.
Shoemaker, Sydney, 1975. “Functionalism and Qualia”. *Philosophical Studies*, 27:292–315.
Shoemaker, Sydney, 1980. “Causality and Properties”. In Peter van Inwagen, editor, *Time and Cause*, 109–35. Dordrecht: D. Reidel.
Shoemaker, Sydney, 1998. “Causal and Metaphysical Necessity”. *Pacific Philosophical Quarterly*, 79:59–77.
Shoemaker, Sydney, 2000/2001. “Realization and Mental Causation”. In *Proceedings of the 20th World Congress in Philosophy*, 23–33. Cambridge, MA: Philosophy Documentation Center. Published in revised form in Gillett and Loewer 2001, 74–98.

Shoemaker, Sydney, 2003. "Realization, Micro-Realization, and Coincidence". *Philosophy and Phenomenological Research*, 67:1–23.

Shoemaker, Sydney, 2007. *Physical Realization*. Oxford: Oxford University Press.

Sider, Theodore, 2009. "Ontological Realism". In David J. Chalmers, David Manley, and Ryan Wasserman, editors, *Metametaphysics: New Essays on the Foundations of Ontology*, 384–423. Oxford: Oxford University Press.

Sider, Theodore, 2011. *Writing the Book of the World*. Oxford: Oxford University Press.

Silberstein, Michael, 2009. "Emergence". In Tim Bayne, Axel Cleeremans, and Patrick Wilken, editors, *The Oxford Companion to Consciousness*, volume 50, 254–7. New York: Oxford University Press.

Silberstein, Michael and J. McGeever, 1999. "The Search for Ontological Emergence". *Philosophical Quarterly*, 50:182–200.

Simons, Peter M., 1994. "Particulars in Particular Clothing: Three Trope Theories of Substance". *Philosophy and Phenomenological Research*, 54:553–75.

Simons, Peter M., 1987. *Parts: A Study in Ontology*, volume 100. New York: Oxford University Press.

Smart, J. J. C., 1958. "Sensations and Brain Processes". *The Philosophical Review*, 68:141–56.

Smith, John A., 2006. *Qualitative Complexity: Ecology, Cognitive Processes and the Re-Emergence of Structures in Post-Humanist Social Theory*. London: Routledge.

Sperry, Roger, 1986. "Discussion: Macro- Versus Micro-Determination". *Philosophy of Science*, 53: 265–70.

Stalnaker, Robert C., 1996. "Varieties of Supervenience". *Philosophical Perspectives*, 10: 221–42.

Stalnaker, Robert C., 1999. *Context and Content*. Oxford: Oxford University Press.

Stephan, Achim, 2002. "Emergentism, irreducibility, and downward causation". *Grazer Philosophische Studien*, 65:77–93.

Steward, Helen, 2015. "What is Determinism". *Flickers of Freedom blog*. https://philosophycommons.typepad.com/flickers_of_freedom/2015/07/what-is-determinism.html. Accessed August 31, 2020.

Stoljar, Daniel, 2007. "Distinctions in Distinction". In Jesper Kallestrup and Jakob Hohwy, editors, *Being Reduced: New Essays on Causation and Explanation in the Special Sciences*, 263–79. Oxford: Oxford University Press.

Stoljar, Daniel, 2010. *Physicalism*. Abingdon: Routledge.

Stoljar, Daniel, 2017. "Physicalism". The Stanford Encyclopedia of Philosophy (Winter 2017 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/win2017/entries/physicalism/. Accessed August 31, 2020.

Strawson, Peter F., 1962. "Freedom and Resentment". *Proceedings of the British Academy*, 48:1–25.

Stump, Eleonore, 1999. "Dust, Determinism, and Frankfurt". *Faith and Philosophy*, 16: 413–22.

Sturgeon, Scott, 1998. "Physicalism and Overdetermination". *Mind*, 107:411–32.

Swinburne, Richard, 2001. *Epistemic Justification*. New York: Oxford University Press.

Swoyer, Chris, 1982. "The Nature of Natural Laws". *Australasian Journal of Philosophy*, 60:203–23.

Tahko, Tuomas E., 2014. "Boring Infinite Descent". *Metaphilosophy*, 45:257–69.

Tahko, Tuomas E., 2018. "Fundamentality". The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/fall2018/entries/fundamentality/. Accessed August 31, 2020.

Taylor, Elanor, 2015. “Collapsing Emergence”. *Philosophical Quarterly*, 65:732–53.
Thomasson, Amie L., 2007. *Ordinary Objects*. New York: Oxford University Press.
Thomasson, Amie L., 2010. “The Controversy Over the Existence of Ordinary Objects”. *Philosophy Compass*, 5:591–601.
Thompson, Evan, 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.
Thompson, Evan and Francisco J. Varela, 2001. “Radical Embodiment: Neural Dynamics and Consciousness”. *Trends in Cognitive Sciences*, 5:418–25.
Tro, Nivaldo J., 2016. *Chemistry: A Molecular Approach*. New York: Pearson.
Tye, Michael, 1990. “Vague Objects”. *Mind*, 99:535–57.
Tye, Michael, 1995. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
Unger, Peter, 1980. “The Problem of the Many”. *Midwest Studies in Philosophy*, 5:411–68.
van Cleve, James, 1990. “Mind-dust or Magic? Panpsychism versus Emergence”. *Philosophical Perspectives*, 4:215–26.
Van Gulick, Robert, 2001. “Reduction, Emergence and Other Recent Options on the Mind/Body Problem: A Philosophic Overview”. *Synthese*, 8:1–34.
van Inwagen, Peter, 1983. *An Essay on Free Will*. Oxford: Clarendon Press.
van Inwagen, Peter, 1990. *Material Beings*. Ithaca: Cornell University Press.
Varzi, Achille, 2019. “Mereology”. The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/spr2019/entries/mereology. Accessed August 31, 2020.
Walter, Sven, 2006. “Determinates, Determinables, and Causal Relevance”. *The Canadian Journal of Philosophy*, 37:217–43.
Walter, Sven, 2010. “Taking Realization Seriously: No Cure for Epiphobia”. *Philosophical Studies*, 151:207–26.
Wandell, Brian, 1993. “Color Appearance: The Effects of Illumination and Spatial Pattern”. *Proceedings of the National Academy of Sciences*, 90:9778–84.
Wasserman, Ryan, 2010. “Material Constitution”. The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/fall2018/entries/material-constitution/. Accessed August 31, 2020.
Watson, Gary, editor, 1982. *Free Will*. Oxford: Oxford University Press.
Wedgwood, Ralph, 2013. “A Priori Bootstrapping”. In Albert Casullo and Joshua Thurow, editors, *The A Priori In Philosophy*, 226–46. New York: Oxford University Press.
Wiggins, David, 2001. *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.
Williamson, Timothy, 1994. *Vagueness*. London: Routledge.
Wilson, Jessica M., 1999. “How Superduper Does a Physicalist Supervenience need to Be?” *The Philosophical Quarterly*, 49:33–52.
Wilson, Jessica M., 2002. *Physicalism, Emergentism, and Fundamental Forces*. Ph.D. thesis, Cornell University, Ithaca, NY.
Wilson, Jessica M., 2002*a*. “Causal Powers, Forces, and Superdupervenience”. *Grazer Philosophische-Studien*, 63:53–78.
Wilson, Jessica M., 2002*b*. “Review of John Perry, *Knowledge, Possibility, and Consciousness*”. *Philosophical Review*, 111:598–601.
Wilson, Jessica M., 2005. “Supervenience-based Formulations of Physicalism”. *Noûs*, 39:426–59.

Wilson, Jessica M., 2006. “On Characterizing the Physical”. *Philosophical Studies*, 131: 61–99.

Wilson, Jessica M., 2007. “Newtonian Forces”. *British Journal for the Philosophy of Science*, 58:173–205.

Wilson, Jessica M., 2009. “Determination, Realization, and Mental Causation”. *Philosophical Studies*, 145:149–69.

Wilson, Jessica M., 2010*a*. “From Constitutional Necessities to Causal Necessities”. In Helen Beebee and Nigel Sabbarton-Leary, editors, *Classifying Nature: The Semantics and Metaphysics of Natural Kinds*, 192–211. New York: Routledge.

Wilson, Jessica M., 2010*b*. “Non-reductive Physicalism and Degrees of Freedom”. *British Journal for the Philosophy of Science*, 61:279–311.

Wilson, Jessica M., 2010*c*. “What Is Hume’s Dictum, and Why Believe It?” *Philosophy and Phenomenological Research*, 80:595–637.

Wilson, Jessica M., 2011*a*. “Much Ado About ’Something’: Critical Notice of Chalmers, Manley, Wasserman, *Metametaphysics*”. *Analysis*, 71:172–88.

Wilson, Jessica M., 2011*b*. “Non-reductive Realization and the Powers-based Subset Strategy”. *British Journal for the Philosophy of Science*, 94:121–54.

Wilson, Jessica M., 2012. “Fundamental Determinables”. *Philosophers’ Imprint*, 12:1–17.

Wilson, Jessica M., 2013*a*. “A Determinable-based Account of Metaphysical Indeterminacy”. *Inquiry*, 56:359–85.

Wilson, Jessica M., 2013*b*. “Nonlinearity and Metaphysical Emergence”. In Stephen Mumford and Matthew Tugby, editors, *Metaphysics and Science*, 201–35. Oxford: Oxford University Press.

Wilson, Jessica M., 2014*a*. “No Work for a Theory of Grounding”. *Inquiry*, 57:1–45.

Wilson, Jessica M., 2014b. “Hume’s Dictum and the Asymmetry of Counterfactual Dependence”. In Alastair Wilson, editor, *Chance and Temporal Asymmetry*, 258–79. Oxford: Oxford University Press.

Wilson, Jessica M., 2015*a*. “Hume’s Dictum and Metaphysical Modality: Lewis’s Combinatorialism”. In Barry M. Loewer and Jonathan Schaffer, editors, *The Blackwell Companion to David Lewis*, 138–58. Chichester: Blackwell.

Wilson, Jessica M., 2015*b*. “Metaphysical Emergence: Weak and Strong”. In Tomasz Bigaj and Christian Wüthrich, editors, *Metaphysical Emergence in Contemporary Physics; Poznan Studies in the Philosophy of the Sciences and the Humanities*, 251–306. Brill: Leiden.

Wilson, Jessica M., 2016*a*. “Are There Indeterminate States of Affairs? Yes”. In Elizabeth Barnes, editor, *Current Controversies in Metaphysics*, 105–125. New York: Routledge.

Wilson, Jessica M., 2016*b*. “The Question of Metaphysics”. *The Philosophers’ Magazine*, 74:90–6.

Wilson, Jessica M., 2016*c*. “Three Barriers to Philosophical Progress”. In Damien Broderick and Russell Blackford, editors, *Philosophy’s Future: The Problem of Philosophical Progress*, 91–104. Hoboken, NJ: Wiley-Blackwell.

Wilson, Jessica M., 2016*d*. “The Unity and Priority Arguments for Grounding”. In Ken Aizawa and Carl Gillet, editors, *Scientific Composition and Metaphysical Ground*, 171–204. London: Palgrave-Macmillan.

Wilson, Jessica M., 2017. “Determinables and Determinates”. The Stanford Encyclopedia of Philosophy (Spring 2017 Edition), Edward N. Zalta, editor. https://plato.stanford.edu/archives/spr2017/entries/determinate-determinables/. Accessed August 31, 2020.

Wilson, Jessica M., 2018. "Grounding-based Formulations of Physicalism". *Topoi*, 37: 495–512. (First published online in 2016 at https://doi.org/10.1007/s11245-016-9435-7).

Wilson, Jessica M., 2019. "Between Scientism and Abstractionism in the Metaphysics of Emergence". In Sophie C. Gibb, Robin Findlay Hendry, and Tom Lancaster, editors, *The Routledge Handbook of Emergence*, 157–76. Oxford: Routledge.

Wilson, Jessica M., 2020. "Essence and Dependence". In Mircea Dumitru, editor, *Metaphysics, Meaning and Modality: Themes from Kit Fine*, 283–300. Oxford: Oxford University Press.

Wilson, Jessica M., in progress*a*. "On the Notion of Diachronic Emergence".

Wilson, Jessica M., in progress*b*. "Primitivist Fundamentality".

Wilson, Mark, 1982. "Predicate Meets Property". *The Philosophical Review*, 91: 549–89.

Wilson, Mark, 2006. *Wandering Significance: An Essay on Conceptual Behaviour.* Oxford: Clarendon Press.

Wimsatt, William C., 1994. "The Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets". *Canadian Journal of Philosophy*, 24:207–74.

Wimsatt, William C., 1996. "Aggregativity: Reductive Heuristics for Finding Emergence". *Philosophy of Science*, 64:372–84.

Wimsatt, William C., 2007. "On Building Reliable Pictures with Unreliable Data: An Evolutionary and Developmental Coda for the New Systems Biology". In Fred C. Boogerd, Frank J. Bruggeman, Jan-Hendrik S. Hofmeyr, and Hans V. Westerhoff, editors, *Systems Biology: Philosophical Foundations*, 103–20. Amsterdam: Elsevier.

Witmer, D. Gene, 2003. "Functionalism and Causal Exclusion". *Pacific Philosophical Quarterly*, 84:198–214.

Wolff, Johanna, 2015. "Spin as a Determinable". *Topoi*, 34:379–86.

Worley, Sara, 1997. "Determination and Mental Causation". *Erkenntnis*, 46:281–304.

Wyszecki, G. and W. S. Styles, 1982. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. New York: John Wiley and Sons, Inc., second edition.

Yablo, Stephen, 1992. "Mental Causation". *The Philosophical Review*, 101:245–80.

Yates, David, 2012. "Functionalism and the Metaphysics of Causal Exclusion". *Philosophers' Imprint*, 12:1–25.

Yates, David, 2016. "Demystifying Emergence". *Ergo: An Open Access Journal of Philosophy*, 3:809–41.

Zimmerman, Dean W., 1995. "Theories of Masses and Problems of Constitution". *Philosophical Review*, 104:53–110.

Index of Names

Subject Index